HYPOTHESIS DRIVEN ASSESSMENT
OF AN NMR CURRICULUM

A Dissertation in
Chemistry
by
Kimberly Cossey

© 2008 Kimberly Cossey

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2008
The dissertation of Kimberly Cossey was reviewed and approved* by the following:

Karl T. Mueller  
Professor of Chemistry  
Dissertation Advisor  
Chair of Committee

Thomas Mallouk  
DuPont Professor of Materials Chemistry and Physics

Mark Maroncelli  
Professor of Chemistry

Rayne Sperling  
Associate Professor of Education, Professor-in-Charge

Ayusman Sen  
Professor of Chemistry  
Head of the Department of Chemistry

*Signatures are on file in the Graduate School
Abstract

The goal of this project was to develop a battery of assessments to evaluate an undergraduate NMR curriculum at Penn State University. As a chemical education project, we sought to approach the problem of curriculum assessment from a scientific perspective, while remaining grounded in the education research literature and practices. We chose the phrase **hypothesis driven assessment** to convey this process of relating the scientific method to the study of educational methods, modules, and curricula. We began from a hypothesis, that deeper understanding of one particular analytical technique (NMR) will increase undergraduate students’ abilities to solve chemical problems. We designed an experiment to investigate this hypothesis, and data collected were analyzed and interpreted in light of the hypothesis and several related research questions.

The expansion of the NMR curriculum at Penn State was funded through the NSF’s Course, Curriculum, and Laboratory Improvement (CCLI) program, and assessment was required. The goal of this project, as stated in the grant proposal, was to provide NMR content in greater depth by integrating NMR modules throughout the curriculum in physical chemistry, instrumental, and organic chemistry laboratory courses. *Hands-on contact with the NMR spectrometer and NMR data and repeated exposure of the analytical technique within different contexts* (courses) were unique factors of this curriculum. Therefore, we maintained a focus on these aspects throughout the evaluation process.

The most challenging and time-consuming aspect of any assessment is the development of testing instruments and methods to provide useful data. After key
variables were defined, testing instruments were designed to measure these variables based on educational literature (Chapter 2). The primary variables measured in this assessment were: depth of understanding of NMR, basic NMR knowledge, problem solving skills (HETCOR problem), confidence for skills used in class (within the hands-on NMR modules), confidence for NMR tasks (not practiced), and confidence for general science tasks. Detailed discussion of the instruments, testing methods and experimental design used in this assessment are provided (Chapter 3).

All data were analyzed quantitatively using methods adapted from the educational literature (Chapter 4). Data were analyzed and the descriptive statistics, independent t-tests between the experimental and control groups, and correlation statistics were calculated for each variable. In addition, for those variables included on the pretest, dependent t-tests between pretest and posttest scores were also calculated.

The results of study 1 and study 2 were used to draw conclusions based on the hypothesis and research questions proposed in this work (Chapter 4).

_Data collected in this assessment were used to answer the following research questions:_

1. Primary research question: Is depth of understanding of NMR linked to problem solving skills?
2. Are the NMR modules working as intended? Do they promote depth of understanding of NMR?
   a. Will students who complete NMR modules have a greater depth of understanding of NMR than students who do not complete the modules?
   b. Is depth of understanding increasing over the course of the experiment?
3. Is confidence an intermediary between depth of understanding and problem solving skills? Is it linked to both variables?
4. What levels of confidence are affected by the NMR modules?
   a. Will confidence for the NMR class skills used in the modules themselves be greater for those who have completed the modules?
   b. Will confidence for NMR tasks not practiced in the course be affected?
   c. Will confidence for general science tasks be affected?
   d. Are different levels of confidence (class skills, NMR tasks, general science tasks) linked to each other?

   Results from this NMR curriculum assessment could also have implications outside of the courses studied, and so there is potential to impact the chemical education community (section 5.2.1). In addition to providing reliable testing instruments/measures that could be used outside the university, the results of this research contribute to the study of problem solving in chemistry, learner characteristics within the context of chemical education studies, and NMR specific educational evaluations. Valuable information was gathered through the current method of evaluation for the NMR curriculum. However, improvements could be made to the existing assessment, and an alternate assessment that could supplement the information found in this study has been proposed (Chapter 5).
# Table of Contents

List of Figures ........................................................................................................... ix

## Chapter 1. The Problem: Why Evaluate?

### Why are We Evaluating? ......................................................................................... 1

1.1 Significance of This Research ................................................................. 1
   1.1.1 The Contribution of Curriculum Assessment to Chemical Education ......................... 1
   1.1.2 Contribution of Penn State NMR Curriculum to Chemical Education .............................. 6

1.2 Related Curricula: The Current State of NMR in Undergraduate Chemistry Curricula ............................... 9

1.3 Why More NMR in the Undergraduate Curriculum? ............................................. 15

1.4 Modules Created for the PSU Undergraduate Curriculum ....................................... 16
   1.4.1 Chem 431W: Advanced Organic and Inorganic Synthesis ........................................... 19
   1.4.2 Chem 457: Physical Chemistry Laboratory ................................................................. 27
   1.4.3 Chem 426: Chemical Spectroscopy .............................................................................. 32

## Chapter 2. Evaluation: How Can We Measure Success of the Curriculum? ................. 38

2.1 Previous Assessments for Chemistry Curricula ..................................................... 38

2.2 Variables of Interest and Hypothesis .................................................................... 41
   2.2.1 Exposure to Hands-on NMR Modules ........................................................................ 45
   2.2.2 Depth of Understanding ......................................................................................... 47
   2.2.3 Problem Solving .................................................................................................... 52
   2.2.4 Confidence for NMR Tasks .................................................................................... 54
   2.2.5 Hypothesis and Research Questions ....................................................................... 55

2.3 Measuring the Constructs: Instruments Used and Developed ................................. 58
   2.3.1 Exposure to Hands-on NMR Modules ........................................................................ 59
   2.3.2 Depth of Understanding and Basic NMR Knowledge .................................................. 63
   2.3.3 Problem Solving .................................................................................................... 67
   2.3.4 Confidence ............................................................................................................ 72

## Chapter 3. Methods Used for Testing, Scoring and Data Analysis ................................. 81

3.1 Testing Methods and Experimental Design .......................................................... 81
   3.1.1 Online Testing ..................................................................................................... 81
   3.1.2 Experimental Design ............................................................................................ 86
3.1.3 IRB Approval.................................................................93
3.2 Pilot (Fall 2005 · Spring 2006)..................................................95
3.3 Data Collection Methods for Studies 1 and 2
(Fall 2006 · Spring 2007)..............................................................99
  3.3.1 Participants .................................................................99
  3.3.2 Instruments .................................................................101
  3.3.3 Procedures .................................................................102
3.4 Rubrics and Scoring...........................................................102
  3.4.1 Basic NMR Knowledge Rubric...........................................103
  3.4.2 Depth of Understanding Rubric.........................................104
  3.4.3 Problem Solving Rubric..................................................111
3.5 Reliability Analysis ..........................................................120
  3.5.1 Interrater Reliability & Scorer Training..............................120
  3.5.2 Item analysis ...............................................................126
3.6 Data Analysis Methods .....................................................129
  3.6.1 Descriptive Statistics and Normality .................................130
  3.6.2 Independent Samples t-Tests ..........................................133
  3.6.3 Pearson’s Correlations ..................................................135
  3.6.4 Comparison of Pretest and Posttest Scores.........................136
  3.6.5 Power Analysis ............................................................137

Chapter 4. Results and Discussion .............................................139
4.1 Study 1: Fall 2006, Chem 457 and Chem 431 .........................139
  4.1.1 Descriptive Statistics and Normality ..............................139
  4.1.2 Independent Samples t-Tests .......................................142
  4.1.3 Pearson’s Correlations ................................................145
  4.1.4 Comparison of Pretest and Posttest Scores ......................152
4.2 Study 2: Spring 2007, Chem 457 ........................................154
  4.2.1 Descriptive Statistics and Normality .............................154
  4.2.2 Independent Samples t-Tests .......................................157
  4.2.3 Pearson’s Correlations ................................................161
  4.2.4 Comparison of Pretest and Posttest Scores ......................165
4.3 Summary of Results and Further Discussion .........................168

Chapter 5. Program Evaluation: Assessment
from Another Perspective .........................................................181
5.1 What is Program Evaluation? .............................................181
5.2 Strengths and Weaknesses of the Previous Method of NMR
  Curriculum Assessment .........................................................184
  5.2.1 Strengths of the Previous Assessment .............................184
List of Figures

Figure 1.1. Laboratory Courses in the PSU Chemistry Curriculum That Contain NMR Experiments .......................................................... 18

Figure 1.2. Schedule of NMR Related Training/Assignments in Chem 431W. ...................................................................................... 22

Figure 1.3. Unique Pedagogical Features of the Chem 431 Tutorial. ................................................................................................. 24

Figure 1.4. Inversion Recovery $T_1$ Experiment ........................................ 31

Figure 1.5. Student Building an NMR Probe for Special Project in Chem 426 ............................................................................. 35

Figure 1.6. Student-Built NMR Probe: Tuning and a Spectrum Produced. ......................................................................................... 37

Figure 2.1. Variables Chosen for NMR Curriculum Study. ....................... 42

Figure 2.2. Hypothesis Used to Assess NMR Curriculum.......................... 56

Figure 2.3. Variables as Defined by Researchers and Assessment ................ 59

Figure 2.4. Depth of Understanding Items .................................................. 65

Figure 2.5. Item Summary for Basic NMR Knowledge Section.................... 67

Figure 2.6. HETCOR Spectrum Used in Assessment .................................. 70

Figure 2.7. Task Confidence Items for General Science and NMR Tasks. ....................................................................................... 77

Figure 2.8. Task Confidence Items for Class Skills ....................................... 79

Figure 3.1. Summary of Instruments Used for NMR Curriculum Assessment ................................................................. 82

Figure 3.2. Planned Testing Schedule for NMR Assessment ....................... 88
Chapter 1

The Problem: Why evaluate?

What are We Evaluating?

1.1 Significance of This Research

1.1.1 The Contribution of Curriculum Assessment to Chemical Education

While curriculum assessments are common in the education realm, professionals in scientific fields have yet to adopt a systematic way of measuring and evaluating their own educational programs and materials. Assessment is the formal term encompasses “all of the processes involved in making decisions” about some product, in this case a new NMR curriculum at Penn State University. Formulating an assessment for a curriculum involves designing measures, tools that aid in gathering information relating to the assessment process, and evaluation, “the process of making decisions on the basis of measurements”, once data is collected. Both chemists and educational researchers have distinct perspectives and can contribute valuable information to better understanding of the processes of teaching and learning chemistry concepts. On one hand, chemists have a thorough understanding of the content area to be taught, and of the processes that make up physical science research. On the other, educational researchers have a deep knowledge of the many student attributes and processes that contribute to learning. They also know which teaching strategies have been shown (through published studies) to be effective for different situations. In the past, there has been
hesitancy from both sides to work together, and a lack of communication has likely hindered both in their search for answers. Chemical education as a field aims to close this gap.

Chemists as a group seem ready to adopt this change. In the field of chemistry, research used to refer only to experiments run in a laboratory within one of the traditional fields of chemistry (organic, inorganic, analytical, or physical chemistry). However, the modern chemist cannot be defined so narrowly. Often research is interdisciplinary and project-based, where a chemist may represent only one part of a larger team of STEM (Science, Technology, Engineering, and Mathematics) professionals. As chemists redefine our contribution to modern problems, we have realized the unique perspective we have to offer the field of education. Not only can we help to better train our own future colleagues (chemists, chemical engineers, etc.), but we have also realized the need to educate the general public on what chemists do, and how our research benefits society. The latter is considered integral to ensuring public funding of government science agencies, such as the National Science Foundation (NSF), in the future. Since government funding to areas such as science is a function of the political climate, scientists, particularly chemists, need to assert their role as improvers of society in order to ensure continued support. Education of the public on the importance of chemistry to their way of life can lead to the election of leaders who will make decisions that positively impact science funding and policy.

Beginning with a few interested professionals in traditional chemical fields, the budding field of chemical education has grown to include professional divisions
(e.g., CHED division of the ACS), peer-reviewed journals (e.g. Journal of Chemical Education), and national and international meetings (e.g. the Biennial Conference on Chemical Education, BCCE). More opportunities for full-time careers in the field of chemical education are cropping up every year. In fact, a recent survey of professional periodicals over six months revealed four advertisements for opportunities for full-time positions such as tenure-track professors (doing only chemical education research) or directors of chemical education outreach programs. However, in order for chemical education as a field to become fully integrated into the chemistry community certain biases must be overcome.

Roadblocks in academia are illustrative of these biases. Faculty who wish to contribute to chemical education research may be penalized in the quest for tenure. The term scholarship, within the chemistry community, has become synonymous with research published in peer review journals. As would be expected, research is a primary component in most tenure decisions for faculty in STEM fields. However, in many chemistry departments, the only research that is acceptable is “hard” research, research completed in a laboratory, preferably within one of the core fields of chemistry (organic, inorganic, analytical, or physical chemistry). Some feel that educational research should be left to the social sciences, areas such as psychology and sociology, and there is a resistance to mixing the two types of research. At some universities, publications in chemical education and educational research journals are not counted towards tenure. If these are not counted as research, then these types of studies would actually be taking time away from compiling a successful tenure package. Not treating educational studies as research discourages new
professors from participating in chemical education. To counteract this attitude, the Chemical Education division of the ACS recently issued the following statement on scholarship.

*ACS Statement on Scholarship, 2006* (excerpt)

Rigorous scholarship in discovery, integration, application, and the study of teaching and learning is needed to foster the innovations that will ensure our economic health in a global economy. By endorsing a broader definition of scholarship, ACS will join other professional organizations in many disciplines, including the arts, humanities, social sciences, mathematics, engineering, and the sciences, in acknowledging the importance of converting information into knowledge that builds upon and informs the work of others...The scholarship of teaching and learning is still perhaps the least understood and recognized of all the forms of scholarship, but has the potential to transform chemical education. It must be encouraged and its role in preparing scientists for the new millennium must be recognized...The chemistry community must accept and act upon a broader definition of scholarship, rewarding faculty for the wide range of activities needed to bring about a modern and effective research and education infrastructure. With original, creative and rigorous scholarly activity that encompasses the entire spectrum of intellectual activities, the chemical sciences and engineering will continue to be at the forefront of innovation.
This statement not only manifests the importance of research into the processes of learning chemistry-related concepts, but also promotes chemists as the directors of this research. In addition, the ACS alludes to the “wide range of activities” that chemical education research can also entail, such as: development of curricular materials, public outreach, and assessment of new programs in the latter two areas.

The ability to formally assess a chemical education project is a vital skill to any faculty wishing to make an impact in the chemical education community. The National Science Foundation has recently recognized the need for assessment of all funded grants, including curriculum and education research. Chemists are required to supply measurable outcomes of a program’s success in order to receive additional funding. Assessment is also important if chemistry faculty hope to disseminate their educational methods to be used at other universities. While conscientious professors often develop a sense for “what works”, their cohorts are unlikely to adopt new methods unless compelling evidence is rendered. For these reasons, chemists require a systematic method of assessing their curricula.

For this study, we use the phrase **hypothesis driven assessment** to convey our process of relating the scientific method to the study of educational problems. This project arose from a discerned need for advanced analytical instrumentation to play a greater role in undergraduate chemistry curriculum. As will be discussed further in section 1.1.2, Nuclear Magnetic Resonance (NMR) became the focus of a series of hands-on modules to be added into the curriculum at Penn State University. Thus, the research question was one that is common in the educational realm: is the new curriculum working? A search of relevant literature
included research into what other chemical educators have done to assess similar programs, the current state of NMR in the undergraduate laboratory courses, and what insights educational studies might have to offer into the learning of an advanced technique such as NMR. We then sought to find the relevant variables and posited relationships among them in a hypothesis. When designing experimental procedures to test our hypothesis, methods commonly used in educational research were used as models. Data analysis was performed using quantitative (statistical analysis) and descriptive methods. Conclusions based upon these data were used to support or disprove the hypothesis, and could influence whether or not the program is continued.

1.1.2 Contribution of Penn State NMR Curriculum to Chemical Education

The NMR curriculum project at PSU arose from a Course, Curriculum, and Laboratory Instruction (CCLI) grant funded through the Department of Undergraduate Education (DUE) of the National Science Foundation (NSF). Primary investigator, Dr. Karl Mueller, and several other professors and instructors at Penn State, sought to improve the laboratory curriculum in the Department of Chemistry by increasing the presence of advanced analytical techniques in undergraduate courses. Nuclear Magnetic Resonance (NMR) was chosen as the principal technique due to its complexity, availability (a new 400 MHz NMR spectrometer had been recently purchased for the undergraduate labs through a separate NSF grant), and the expertise available from faculty and
facilities within the chemistry department at the university.

The goal of the CCLI grant was to “develop a comprehensive suite of NMR laboratory exercises that can be implemented within the physical chemistry laboratory or vertically integrated into the organic, physical, and instrumental analysis courses” at Penn State. Why were additional NMR modules required? According to the grant’s authors and a previous literature search, NMR was rarely being used to illuminate key concepts in physical chemistry even though the technique itself is built upon theory from this field. Numerous laboratory experiments have been published that utilize NMR for organic and structural applications, but by revealing only one aspect of this complex technique, “the educational potential of one of the most elegant and sophisticated tools for probing chemical systems is never fully achieved and is at risk of being nothing more than an expensive data logger.”

This curriculum aims to cover NMR in greater depth and to foster the connection of NMR concepts within and across courses. Experts in any content area are defined by their ability to make connections among previous knowledge and new information. In modern undergraduate chemistry curricula, the pressure to cover all of the areas that such a broad field entails often produces graduates who are experts in nothing. While this breadth of coverage is important, sacrificing depth of understanding could also mean depriving students of the opportunity to practice linking knowledge together. In fact, “research on expertise suggests that a superficial coverage of many topics in a domain may be a poor way to help students develop” skills that they will need in their future careers. We expect science majors
to graduate with the ability to solve problems, but this skill is based on the ability to choose and connect relevant information. This project sought to increase students’ depth of understanding in a particular content area (NMR) by introducing the multiple aspects of this complex instrument (applications, theory, and instrumentation) and revealing the connections among them.

When deciding on an analytical technique that warranted in-depth treatment, NMR was an ideal choice because it is a modern analytical technique that can be integrated into multiple areas of chemistry. Realizing that there are many other important chemistry topics still to be learned by the undergraduates, researchers sought to reduce the impact to any one area of chemistry by spreading NMR modules across curriculum, introducing the relevant aspects of the technique in organic, physical, and analytical laboratory courses. In addition, it was proposed that students would gain a greater appreciation for the intricacies and utility of modern analytical instrumentation through their hands-on experience\(^1\) with the NMR spectrometer.

The NSF requires all those receiving funding for CCLI grants to submit an assessment of the project, and fulfillment of this requirement lead to the Ph.D. work presented in this thesis. It is not enough to produce new curricular materials. It must also be determined what effect, if any, these new NMR modules have on the students taking laboratory courses.

Through discussions with the grant’s authors, goals were clarified and expected student outcomes were put into more concrete terms. By covering more complex information in one analytical technique (NMR), the authors wanted
students to learn:

1. To synthesize complex information into a more cohesive view of NMR, which would require students to make connections within and among the various fields of knowledge within NMR (theory, application, and instrumentation).

2. To move past the novice’s perspective of NMR, and view the technique more as an expert (in NMR) would.

3. To translate gains (in knowledge, skills, confidence, etc.) from advanced experiences in one specific area (NMR) to future experiences with other analytical techniques.

Overall, it is expected that achieving the above goals would better prepare undergraduates for future careers as professionals in chemistry-related fields. For the purposes of this assessment, goals 1 and 2 were grouped under the term depth of understanding. Goal 3 seeks to link outcomes gained through one technique to those for another technique. In order for learning in the area of NMR to affect other knowledge areas, transfer of knowledge must occur. These goals were used in designing assessing tools to measure and evaluate related students’ skills (e.g. problem-solving) and attributes (e.g. confidence) for the task(s) being studied.

1.2 Related Curricula: The Current State of NMR in Undergraduate Chemistry Curricula

NMR is an ideal instrument to examine dynamic molecular processes. Determining the rotational energy barrier of an amide bond using analysis of NMR
line-shapes at multiple temperatures is a classical experiment for physical chemistry.\textsuperscript{11} However, these examples are rare in undergraduate chemistry curricula and are still being improved upon. Morris and Erickson\textsuperscript{12} have recently reported a modification that adds saturation transfer techniques (and necessary T\textsubscript{1} determination) to improve students’ calculations of enthalpy and entropy of activation. In a more simplified study, Weil\textsuperscript{13} has found that bond rotation of a hydrogen-bonded picryl system can also be used to show the conformational exchange of protons through line-shape analysis at various temperatures and T\textsubscript{2} determination. Electron exchanges can be monitored by NMR as well, and this has been shown by Jameson.\textsuperscript{14} The fast-exchange process between a diamagnetic species (ferrocene) which oxidizes to form a paramagnetic species (ferrocenium ion) can be monitored by relating shifts and line-widths of peaks from each species to determine the mole fraction of each.

Cutting-edge NMR research looks at systems that extend far beyond the standard liquid samples examined in most undergraduate laboratories. Unfortunately, only a few procedures that reflect utilize NMR spectrometers in this manner can be found in the literature. Solid-state NMR is an invaluable analysis technique for material sciences. In an experiment reported by Anderson,\textsuperscript{15} students synthesize sodium phosphate glasses and analyze percent composition through \textsuperscript{31}P NMR, and then attempting to determine the glass compositions by comparing their spectra to computer-simulated spectra of various compositions. Magnetic Resonance Imaging (MRI) is another fascinating application of NMR, and Quist has developed an imaging experiment that uses can be used on a 100 MHz NMR spectrometer.\textsuperscript{16}
The sample consists of cylinders of water, made by filling cylindrical holes in a poly(tetrafluoroethylene) plug which is then placed in a normal NMR tube. After applying gradients, created using the X, Y, and Z shims on the spectrometer, the image is reconstructed through back-projection of the spectra.

NMR has been used extensively to aid in structure determination of organic molecules, and this is the most common use of the instrument in the undergraduate lab. Indeed countless experiments using NMR to determine organic structures can be found in the literature. What is less common are those experiments that show this aspect, but then extend students’ knowledge a step further. Several recent examples include conformational analysis of brominated cyclohexanone products,\(^{17}\) and a discovery approach to the concept of shift additivity using a series of aromatic compounds.\(^{18}\) One especially interesting example for novice level students is a guided inquiry introduction to NMR. Students first compare a series of related compounds that increase in number of peaks (acetone, acetic acid, etc.). Then they use this information to pick out the product and by-products of an aspirin synthesis.\(^{19}\) In addition, there is an increase in the number of organic projects that relate to other fields such as inorganic and biochemistry. For example, synthesis of organometallic compounds presents an opportunity for students to use multinuclear NMR.\(^{20}\) In another experiment, Peterman, et al. has students conduct an enzyme study using \(^{19}\)F NMR.\(^{21}\)

Physical organic experiments reveal another interesting application of NMR that can show how molecules behave. For example, if scientists want to explore the mechanism of a reaction, they can perform an isotope effect study. Adapted for the
undergraduate laboratory by Chechik, the reaction rates of both deuteration and bromination of hexanone are measured. By comparing the reaction rates, it can be determined which bond is being broken in the rate-limiting step. Spectroscopy can also be used to gain information on bonding. Mosher has developed an experiment that demonstrates the different orbital mixing available to carbon atoms bonding in an organic molecule. In this experiment, the $J$-coupling between C and H is related to percent $s$ character in example molecules. Once the “standard” $J$ values for sp, sp$^2$, and sp$^3$ are determined, they can be compared to other non-standard C-H bonds.

Since the work on the NMR project at PSU has begun, a collection of innovative NMR experiments and approaches to improving NMR education in the undergraduate setting has been published by the ACS. “Modern NMR Spectroscopy in Education” not only contains novel experiments developed for specific courses in the usual chemistry curriculum, but also discusses implementation issues and contains a special section with strategies for those with lower field, permanent magnet spectrometers. Of particular relevance to this work are Fisher and Fish's efforts to develop an undergraduate chemistry curriculum in which NMR serves as “a unifying and integrative thread.” NMR activities have been added to the general chemistry, organic chemistry and biochemistry courses at St. Vincent College. The majority of additions are pen-and-pencil exercises such as correlating electronegativity of neighboring groups with H-NMR shift (general chemistry) or peak assignments for a variety of 2D spectra (organic and biochemistry). The most obvious difference between this and our work in NMR education is that no hands-on experiments have yet been incorporated into St.
Vincent’s curriculum, though there are plans to add a labeling experiment and a kinetics experiment in the future. The primary focus of Fisher and Fish’s assessment appears to be gains in specific NMR skills. An assessment of graduating seniors has shown an increase in the percent of students who can assign structures correctly using NMR data. In addition, longitudinal studies show that graduating seniors are better at determining a structure using NMR data than they were when they were second semester sophomores. However, improvement over the four year degree may be a result of skills being practiced over time or students maturing, rather than a direct result of new additions to the curriculum. The focus of this assessment on structure determination as a primary outcome also differs fundamentally from the goals of the PSU curriculum project (which seeks to broaden students’ views of NMR’s utility). It will be interesting to see if students are making gains in other areas, and though the exact assessment plan is not presented, authors are considering using “primary trait analysis”\textsuperscript{26,27} and possibly Nuhfer and Knipp’s\textsuperscript{28} Knowledge Survey.

Also relevant to this work are studies that highlight the importance of hands-on use of the NMR spectrometer. At Pomona College, NMR techniques are highlighted in the physical chemistry, capstone laboratory course.\textsuperscript{29} As stated within this publication, one goal of the course was for students to “have hands-on experience with research-level instrumentation.” In addition, these experiments expanded students’ exposure to experiments that give chemical information beyond structure, which is one of the goals of the Penn State curriculum. Not only did students at Pomona College complete the previously mentioned dynamics
experiment, which analyzes the barrier of rotation for an amide, but they also performed a relaxation experiment very similar to that chosen for our project. In the latter, relaxation constants $T_1$ and $T_2$ are measured. Students entered their own set of delay times, and relaxation constants were used to obtain a rotational correlation time (measure of rotational mobility). In addition, the authors are currently developing an MRI experiment that is straightforward and executable for a typical NMR spectrometer with gradients. Hanson has developed a protocol that allows sophomore organic students to access the NMR spectrometer through automation, anytime of the day or night. This method was in contrast to the previous strategy where TAs ran all NMR experiments, and students “either watched the process passively or went back to lab.” Through guidance from TAs and professors, students are encouraged to run additional NMR experiments to obtain more information about their unknowns or reaction products. The utility of these experiments is explained only after they are run, allowing new concepts to be introduced stepwise and when they are of the most relevance to the students’ work. While there were no formal measurements of whether this approach improved students knowledge of NMR, the author reports that many students were “highly motivated to learn more about it” when NMR was introduced in lecture, weeks after its initial use in the laboratory. He also suggests that students have “learned that the NMR technique answers the most practical of questions: what happened?” While the intention of using automation was to allow to students have the freedom to direct their experiments (choosing pertinent additional experiments), it is unclear whether students took the initiative to do so (after they knew of additional experiments) or
whether they were continually directed by their mentors.

1.3 Why More NMR in the Undergraduate Curriculum?

The most unique feature of the new NMR curriculum developed at Penn State was its focus on integrating one analytical technique (NMR) across multiple courses, in order to cover the many complex aspects of this instrument. However, addition of new materials will require some older materials to be replaced. Any change to a chemistry curriculum must be justified, if it is to be adopted permanently (or to be exported to other universities). Thus, we must answer the question, what are the benefits of a more in-depth treatment of NMR?

Besides supplying undergraduates with the knowledge they need in their future careers, courses and laboratories should give them an idea of what their chosen profession entails (daily activities, skills required, etc.)\(^{10}\) Analytical instruments perform many sophisticated techniques that have become indispensable to researchers in chemistry and it should be a goal of undergraduate chemistry education to incorporate some of these techniques into the laboratory curriculum. Chemistry students are introduced to many scientific instruments throughout their undergraduate careers, but often depth of understanding is sacrificed for breadth of exposure.\(^{8}\) Showing only one application of an analytical instrument conveys the misconception that this expensive equipment is only useful for the most obvious and simple applications. A professional chemist must be able to synthesize many aspects of a complex problem in order to arrive at a valid conclusion. The separation of chemistry courses can lead students to isolate information learned in what they
deem to be different areas. Thus when they choose their particular field of interest, some information is disregarded. This problem could be mitigated by integrating aspects of a complex technique and distributing it across the chemistry curriculum.

The chosen analytical instrument must be easily adapted to span multiple courses and should be sufficiently complex to warrant an in-depth treatment: NMR meets both of these criteria. Even when treated in multiple courses, analysis of NMR data is usually limited to structural information while a wealth of additional applications for NMR are left unexplored (e.g., dynamics, kinetics, etc.). By distributing diverse materials, designed to best fit the context of each course across the curriculum, the complex phenomenon of NMR can be approached from many different directions and students can make the necessary connections. In addition, more time can be devoted to discussing complicated concepts to ensure a deeper understanding. Cover one topic in greater depth does mean sacrificing some time that would be spent on other techniques, however, details of a less familiar technique can always be picked up when needed (as is often expected in industry positions). In fact, an appreciation for the power of an analytical instrument to solve chemical problems, shown through this curriculum, could provide the necessary motivation and self-confidence to tackle the new instruments.

1.4 Modules Created for the PSU Undergraduate Curriculum

New course materials that cover NMR in greater depth were developed as part of a CCLI grant® funded by the Department of Undergraduate Education
By designing modules that linked a particular aspect of NMR to a course in the chemistry curriculum, researchers intended to communicate a deeper, more comprehensive view of NMR to the students. Figure 1.1 shows the laboratory courses where NMR experiments were run, either as part of the previous curriculum or as part of the new materials added through this project. Course availability and prerequisites determine when students take these courses. It is important to note that students in the Advanced Organic and Inorganic Synthesis course (Chem 431W) are required to complete the basic organic laboratory (Chem 36) as a prerequisite before they can begin work in Chem 431W. Similarly, students cannot enter the Chemical Spectroscopy course (Chem 426) without having first completed the physical chemistry laboratory (Chem 457). New modules were distributed across the chemistry laboratory curriculum by matching the NMR concepts with the course that provided the most appropriate context. Organic courses (such as Chem 36 and 431W) were the most common and obvious place to introduce the primary application of NMR, structure determination (1D and 2D experiments), in greater detail. The physical chemistry laboratory (Chem 457) offered a venue to link the abstract quantum mechanical aspects of NMR theory with hands-on experiments. Specifically, experiments chosen for this laboratory not only illustrated intricate details of NMR theory, but also demonstrated the power of manipulating this information to access additional information about a chemical system (non-structural information, e.g. dynamics). Chemical Spectroscopy (Chem 426) serves as an instrumental analysis course at Penn State, and thus provided an opportunity for students to gain a deeper knowledge of the inner workings of an NMR spectrometer.
All new materials were designed to be modular, meaning that the new NMR content of each course could be taught independently. Students who learned all of the available modules were receiving a comprehensive view of NMR, regardless of the order of courses taken. Researchers felt it was important for students to gain direct experience with the recently purchased 400 MHz NMR spectrometer, and they developed at least one hands-on experiment for each upper level (400 level) course.

**Figure 1.1. Laboratory Courses in the PSU Chemistry Curriculum That Contain NMR Experiments.** The chart below shows the “normal” schedule for a chemistry major taking these lab courses over a 4 year track. Prerequisites for the laboratory courses are also listed below. Those course followed by “(concurrent)” may be taken in the same semester as the lab course.

<table>
<thead>
<tr>
<th>Student Year (4 Year Track)</th>
<th>Semester</th>
<th>Laboratories that contain NMR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Organic (Chem 36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Advanced Org. &amp; Inorg. Synthesis (Chem 431W)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physical Chemistry (Chem 457)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chemical Spectroscopy (Chem 426)</td>
</tr>
<tr>
<td>Sophomore</td>
<td>Fall</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Spring</td>
<td>X</td>
</tr>
<tr>
<td>Junior</td>
<td>Fall</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Spring</td>
<td>X</td>
</tr>
<tr>
<td>Senior</td>
<td>Fall</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Spring</td>
<td>X</td>
</tr>
</tbody>
</table>

**Course:**

**Prerequisite Labs**

Chem 36: No labs listed

Chem 431W: Chem 36

*Chem 457: No labs listed

**Prerequisite Lectures**

Chem 36: Chem 38 and 39 (concurrent) (Organic I & II)

Chem 431W: No lectures listed


Chem 426: Chem 457

Chem 426: Chem 452 (P-Chem: Quantum Mechanics)

* For this course, prerequisite Chem courses may be substituted for Ch E 320 (Phase and Chemical Equilibria, a chemical engineering course).
Though the main focus of this NMR curriculum was hands-on lab modules, lectures and homework assignments were also generated for most courses. One key feature of the study was to promote depth of understanding through multiple exposures to NMR content. For this reason, a brief overview in either lecture (Chem 426 and 457) or written form (Chem 431, within the lab guide) was given to the students early in the semester. Not only does this help those students with less exposure to NMR through other courses to catch up to their cohorts with more experience, but it also provides a model for selecting key concepts and connecting information from another context to the current course.

1.4.1 Chem 431W: Advanced Organic and Inorganic Synthesis

The Course

Chem 431W is an advanced laboratory course that focuses on synthesis techniques and scientific writing. This course integrates aspects of modern chemical research, and can act as a transition from the lower-level content based laboratories into project-based student research. For example, students are required to perform experiments directly from the primary literature because these expose students to non-ideal reactions; where the first attempt to run a reaction may give low yields, requiring conditions to be optimized and products purified. In addition, students must learn to manage time and resources effectively, as many course assignments overlap and there is competition for instrument time. This course requires students to present their work in the style most common in modern chemistry: ACS style reports and poster presentations. Reports are more sophisticated, as they require
motivation for the work and concise explanations of procedures and results, where
students must learn to choose what is relevant to their audience and to build a case
for their conclusions.

It is relevant to note that Chem 431W students are building on previous
exposure to NMR in Chem 36, a sophomore level lab course. In the prerequisite
course, students have had an opportunity to use an older (80 MHz) NMR
instrument, but interactions with the instrument itself is limited to a short list of
commands, including exactly what keys to press, without any explanation of what
the instrument is doing or the motivation for the process. Data analysis in Chem 36
is also limited to either identification of an “unknown” (with a given list of options)
or confirmation of the expected product of the reaction a student has run. For some
experiments, students are asked to take an NMR spectrum, however if the spectrum
produced differs in any way from that of the expected product (i.e. the final product
is not pure), the experiment is repeated. In contrast, the solution to such a problem
in Chem 431W would be to attempt to purify the final material and/or to identify the
impurities shown in the spectrum. Chem 36 students may see the utility of an NMR
experiment, but the power of the instrument to disentangle complex information is
saved for the later course. If Chem 36 is a course designed for novices in organic
chemistry and analytical instrumentation, then Chem 431W aims to create experts
in this particular area of chemistry.
New Hands-on NMR Module

NMR samples in Chem 431W are run in automation and thus students primary interaction with the instrument is through data processing using X-WIN software. A primary component of the hands-on activity for this course was a software tutorial, described as “NMR software training”. In it, students learn how to phase, calibrate, integrate, and use peak picking for an NMR spectrum. Good processing is crucial for analysis of NMR data, especially for samples that contain a mixture of compounds. For each step, the tutorial explains not only “how” but also “why” this processing must be done. Students were given a write-up of the tutorial so that they could work at their own pace, but a TA was also available for any questions.

Though the NMR software training was the only new material generated, it was used in the context of several NMR activities that make up the hands-on NMR module for this course (Figure 1.2). Before NMR software training (the tutorial) was completed, students had received instructions on how to submit samples to be run (in automation) on the 400 MHz NMR spectrometer, and had run at least one sample (week one). Students used the tutorial to “work up” data from a standardized sample in week two. This was immediately followed with an assignment where students were required to use their software skills to “work-up” their own sample and analyze the data to determine the components of an unknown mixture in the “analysis of a crude product” assignment (due week three).
NMR software training was presented in the following manner. Students were required to sign-up for time, so that a TA familiar with the material was present to answer questions. It was made clear to students that this training was mandatory and that it would help them on future assignments in the course. Due to the limited number of computers that share the X-WIN software, only two slots were available at a given time and a third computer was left open in case a student was not finished in the allotted time (45 minutes). Students were given the tutorial materials (paper copy) and were allowed to go through the material at their own pace. They could also ask questions to the TA as needed. Upon completion of the tutorial, the TA checked the final print-outs to be sure that each student had produced a properly phased, integrated, and calibrated spectrum. A record was kept of which students underwent training, and this was compared against the list of participants in the study to ensure that all participants had completed this activity. Students were asked to keep the tutorial, so that they would have a reference for the upcoming “analysis of a crude product” assignment.

The tutorial itself consisted of detailed instructions as well as the motivation
behind each step of the process. After an introduction to the software controls and how to find and open their files, students were asked to open and use the file from a standardized NMR sample. This sample was chosen to give a slightly complex spectrum, illustrating features such as overlapping peaks and complex splitting. Students were not allowed to use their own spectra for the tutorial because previous experience during piloting had shown this to take far too much time (many more overlapping peaks to integrate) and to be very confusing (peaks from more than one organic molecule). Using the standard sample allowed students to learn the steps and function of each process (phasing, integration, etc.) in a stepwise fashion and ensured that all students were given a common starting place. After practicing with a simpler sample, they should use the same process on a more complex sample (the impure mixture for the “analysis of a crude product” assignment). Hints on how to approach a more complex spectrum were given throughout the tutorial for later use.

The X-WIN tutorial for Chem 431W was designed to walk students through the manipulation of output from a $^1$H-NMR experiment, however it also contained several features aimed at increasing student understanding of NMR (shown in Figure 1.3). By explaining why certain steps were necessary when manipulating the output (in relating importance segments), researchers were attempting to increase student motivation and to reinforce process skills. The tutorial was also designed to help students make connections between process skills and how the NMR spectrometer works (including its interaction with samples). This was the goal of addressing a common misconception segments. The developing skills segments were included to help students connect what they learned through this tutorial to skills
that will be required in future assignments.

**Figure 1.3. Unique Pedagogical Features of the Chem 431 Tutorial.** Researchers aimed to create a tutorial that not only provided the steps to process an NMR spectrum, but also sought to increase students' understanding of the technique. Unique features were spread throughout the tutorial and are listed below. The types of information given were grouped into three categories (*relating importance, addressing a common misconception*, and *developing skills*). A succinct description of each instance is also given. However, see Appendix A for the complete discussion presented within the tutorial.
Unique Features of X-WIN Data Processing Tutorial

1. Phasing

   *Relating Importance:* Improper phasing will result in bad integration values.

2. Calibrating & solvents

   *Addressing a common misconception:* You must use a deuterated solvent dissolving sample, not the non-deuterated solvent (found in the lab). The deuterated sample is required because:

   i. Deuterium is invisible to the frequency that the spectrometer is tuned to

   ii. The spectrometer must compare the frequency of the nuclei of interest (e.g. the frequency of hydrogen) to compare to another frequency (that of deuterium) in order to stay tuned.

   *Addressing a common misconception:* The peak we calibrate to is not deuterium, but is trace H-sample. It comes from hydrogen exchange with water. (This explains the splitting of the peak into a pentet, rather than a singlet as would be expected.)

   *Developing Skills:* Picking a solvent if NMR sample is NOT soluble in CDCl₃
3. Integrating

*Addressing Common misconception:* Clarification that an integral tells the ratio of signals contained in the peak, not always an absolute number.

*Developing Skills:* Integrating peaks that overlap

*Developing Skills (for future assignment interpretation):* Using integration to determine the relative quantities of components of a mixture

4. Peak picking

*Relating Importance:* Peak picking is used for determining $J$ values. These must be reported for all products in the experimentals section of final reports.

5. Printing close-ups

*Relating Importance:* Printing close-ups will help students to identify peaks that are unclear on full spectrum. This is especially important when reporting splitting and $J$ values.

6. Special Requirements for $^{13}$C or other non-H spectra

*Developing Skills (for future assignments):* When working up a $^{13}$C (or other non-H) spectra there are several difference: reference peaks have different ppm values and splitting, integration is not needed, and peak picking is used to give the actual shift values in ppm.
Once students had become familiar with the X-WIN software, they used it to complete an assignment entitled “Analysis of a Crude Product.” In the modern organic synthesis lab, when reactions are run, the products obtained are rarely pure. The resulting crude products may contain excess reactants, by-products, and/or solvents. For this assignment, students are given a sample and a list of five reaction schemes, one of which was used to make their sample. These reaction schemes show the structures of all reagents, solvents, and expected products. For each reaction, there is also a work-up/isolation step, which contains additional solvents (contaminants) that may be present in the crude mixture. Students are asked to identify all the components of the sample, and mole percent of each (e.g. percent completion of the reaction). Thin Layer Chromatography (TLC) may be used to estimate the number of components, however NMR is the primary method of determining the identities and ratios of these components.

1.4.2 Chem 457: Physical Chemistry Laboratory

The Course

The physical chemistry laboratory at Penn State, Chem 457, aims to provide experimental evidence of topics within thermodynamics and quantum mechanics that have been presented in the corresponding lecture courses (Chem 451 and Chem 452), and to teach students how to correctly interpret and present quantitative data. In previous years laboratory topics included thermodynamics, kinetics, and quantum mechanics. The corresponding laboratory experiments used to cover these were: bomb calorimetry and phase diagrams (thermodynamics), kinetics of Tiron, and
fluorescence quenching experiments (quantum mechanics). While these classic experiments cover the pedagogical goals of the course, they hardly reflect the focus of modern physical chemistry research and technology. NMR is commonly used in cutting edge research, and the theory behind it combines aspects of quantum and statistical mechanics as well as quantitative data manipulation. For this reason, in-depth treatment of NMR theory was added to the physical chemistry laboratory curriculum.

New materials added to the course needed to match the goals and focus of the course. Physical chemistry has a reputation for being a particularly difficult and abstract content area, thus it was vital to provide concrete illustrations of the related concepts. A hands-on NMR experiment that can link calculations and theory to a practical application was well-suited to this goal. As an upper level course often taken by juniors, this laboratory aimed to increase students’ independence in data analysis and reporting, while still providing structured experiments. Students were expected to find and utilize resources outside of class, rather than depending on the lab guide or TA to lead them through questions and data analysis. While each assignment listed the types of data and calculations necessary, students must decide what methods they will use to obtain the data and the format it will be presented in. This provides a model for recognizing pertinent data and choosing supporting information for students’ work, while paving the way for greater independence in subsequent lab courses (e.g. Chemical Spectroscopy, Chem 426).

To provide a context for the new hands-on NMR experiment, lectures were developed to highlight the relevant quantum theory and its application to chemical
systems (to give a wealth of information beyond structure alone). Experimentalists in the field (Professor Karl Mueller), teaching faculty (Drs. Sykes and Ucak-Astarlioglu), and teaching assistant (Nancy Washton) collaborated to create a series of lectures to that introduced and built upon basic principles of quantum theory and NMR. This topic is rooted in advanced mathematics and so activities were created that allowed students to visualize and manipulate the key equations through Mathematica, computer software package commonly used by scientists and engineers.

New Hands-on NMR Module

The students in Chem 457 have varying levels of experience with NMR from previous courses, but one thing they share is that none have run an advanced experiment without using automation. Automation removes the students from the instrument, and this may contribute to a student’s fear of “breaking” the instrument if he/she was allowed to have any direct control over its function. Students will sometimes use this fear of “breaking” something as an excuse to remain uninvolved in the running of experiments. For this course, students completed a T<sub>1</sub> experiment using a student guide, where they were allowed to type in all of the commands and to even choose variables. As with the 431W hands-on tutorial, instructions explained not only “the how” but also “the why” for each step.

There are a number of advanced NMR experiments, but the T<sub>1</sub> experiment was chosen because it yields information about the molecule that goes deeper than just determining its identity or structure. This is unique in that all NMR
experiments offered to the students, prior to Chem 457, have been aimed at determining structural information only. The assignment itself was a modified version of the experiment published by Gasyna and Jurkiewicz. The goal of this lab assignment was to find $T_1$ relaxation times, a property of the magnetic moment of each carbon atom in hexanol, and to relate it to freedom of rotation for each carbon, a dynamic property.

The first half of this lab assignment required students to run the $T_1$ experiment on the 400 MHz NMR spectrometer using the student guide. Students worked on this experiment in pairs, and could do so at their own pace since the guide contained step-by-step instructions. However, the instructor was nearby at all times to facilitate proper usage of the spectrometer and answer any questions. Some aspects of experimental set-up could not be completed by students, due to time constraints. When they entered the NMR room to begin working on the experiment, the instrument had already been loaded with a degassed sample (70% hexanol in CDCl$_3$), tuned (for both $^{13}$C and $^1$H), locked, and the appropriate 90 times for both nuclei had been determined. Though they were not able to do this portion themselves, the reason for these processes were explained within the laboratory experiment. Students began by opening a new file and reading in a set of shims for their sample. Next students enter a set of predetermined parameters ($^{13}$C and $^1$H 90 times, and the delay between scans), and the purpose of each was explained within the student guide. Next students entered their own values for the variable delay (variable $t$) into what is known as a vdlist. As can be seen in Figure 1.4, the variable $\tau$ (a set of delays between the 180° and 90° pulse of the experiment) is key to finding
the correct T1 value. The final T1 values are found by fitting the data to the curve shown in Figure 1.4, and choosing “proper” values of τ can make the difference between a good or poor fit. Students were allowed to pick their own set of values for this variable, and were encouraged to think about the relation of these values to the data they were trying to obtain. After the vdlist was entered, students entered the commands to start the experiment.

**Figure 1.4. Inversion Recovery T1 Experiment.** In the Chem 457 hands-on NMR experiment, an inversion recovery T1 experiment was run by students. The pulse sequence for this experiment is shown below. Students input a series of values for τ (the variable delay between pulses), and the resultant magnetization was used to calculate T1 by fitting data to the equation shown.

If set-up properly, the 400 MHz spectrometer takes about 35 minutes to run the T1 experiment, and after this is complete students must process the data. The foci of this assignment were how an NMR experiment is set-up and what happens during the experiment itself, thus the subsequent processing was not elaborated
upon. Step-by-step instructions for processing the data were given in the student guide that resulted in a pseudo-2D output. When processing was complete, the students had T₁ values calculated by the X-WIN program and also a printout of the raw data used to obtain these values.

Data analysis and interpretation are the most important components of an experiment in this laboratory course, and so this section was used to obtain students to think more deeply about processes within the NMR spectrometer and how the instrument interacts with chemical species to give valuable information. Since the X-WIN program calculates the T₁ values for the students, they are asked to fit the raw data using another program for comparison. After converting T₁ values to τ_c, the rotational correlation coefficient, students can relate this value derived from the NMR experiment to the dynamics of the hexanol molecules. Though the equation to convert the variables is given, students must think through what these numbers mean with regard to the hexanol molecule on their own. A molecular model of hexanol is available to help them visualize and hopefully make connections. This experiment not only reinforces laboratory goals, but also shows how quantum mechanics can be applied to a chemical system to really “see” what the molecule is doing.

1.4.3 Chem 426: Chemical Spectroscopy

The Course

Chemical Spectroscopy (Chem 426) is an elective laboratory course offered to juniors and seniors at Penn State. The stated goal of this course is to “review
modern methods and instruments of spectroscopy and their applications to problems of chemical structure and analysis.” In this course, students learn about instrument design, advantages and limitations of each technique for collecting chemical data, and the underlying principles of each experiment. In order for a student to understand how a specific area of spectroscopy works, they must understand not only the electronic components, but also how to modify parameters to get the desired results for a particular experiment. This is part of an overreaching goal of the course, to train chemistry majors to think through the decisions necessary to set-up their own experiments based on research questions.

In order to achieve high-order thinking of analytical problems, the instructor (Dr. Dan Sykes) organized the course to build-up information in a logical progression. The course begins by introducing electronics and optics, topics that will act as a foundation for all instrumentation. Next, interactions at the atomic level, and then molecular level were described through the context of the relevant chemical spectroscopies. Full-detail of instruments of increasing complexity could then be covered. The techniques covered in this course are: UV-VIS, fluorescence, FTIR, Raman, liquid-and solid-state NMR spectroscopies. However, content knowledge of existing spectroscopy techniques was not the terminal point. Instead, it was proposed that these upper-level science majors would be able to apply what they have learned in novel situations, and this was measured within a project assignment, which each student was required to complete.

Chem 426 offers students the opportunity to build instrumentation for a specific chemical problem, and this project component is what sets this course apart
from other laboratories in the chemistry curriculum. Knowing when to use an instrument is important, but the professor also expected students to learn how to “select appropriate components to build an instrument for specific uses.” Projects were assigned to groups of students (3-5 per group) at the beginning of the semester. This hands-on assignment represented a large portion of each students grade for the course (20%), and consisted of building instrumentation, meeting project benchmarks, written reports, and a final poster presentation.

New Hands-on NMR Module

Few undergraduates can say that they have seen the inner-workings of an instrument as intricate as an NMR spectrometer, however in this course most students will see it and a few will even construct the parts crucial to running NMR experiments (Figure 1.5). The heart of an NMR spectrometer is the probe inside, the surrounding bulk contains the superconducting magnet and layers of cooling liquids (N₂ and He). Probes are designed to hold the sample in the ideal location within the magnet, apply the rf pulses, and spin the sample. These can be removed from the spectrometer, and probes are interchanged depending on the sample and experimental conditions required. In liquid-state experiments, there is rarely a need to change the probe, and so most undergraduates remain unaware of the role different probes can play in an experiment.
Figure 1.5. Student Building an NMR Probe for Special Project in Chem 426. In the picture below, a student is soldering part of an NMR probe. This probe was designed by a project team (of 4 students) in Chem 426. When completed it was placed within the 400 MHz spectrometer and used to run NMR experiments on deuterium-labeled samples.

The introduction of solid-state NMR requires an elaborate set-up (changing probes and cabling) and so represents an excellent opportunity to show students what components normally lie hidden inside the spectrometer. All students are given the opportunity to watch this process, and run a solid-state experiment. NMR discussion takes place during the unit on molecular interactions, so students have already studied the basics of electronics and are familiar with how spectroscopic processes affect matter on the atomic level. In addition, they have had several lectures covering NMR theory (vector and product operator models) and special considerations necessary for solid-state instrumentation. By comparing and
contrasting a liquid- and solid-state NMR probe, students can begin to understand the way that scientists modify instruments to meet their needs. For example, solid samples are conventionally spun at an angle (54.74°) to improve resolution of the peaks in a process called magic angle spinning (MAS). This required experimentalists to develop a new type of probe that had the necessary spinning capabilities and also kept the magnetic fields at the proper orientation.

If seeing the components within the NMR spectrometer is helpful to students’ understanding of instrumentation, then actual building of a working a probe could be considered the pinnacle of this type of learning. As part of the student projects previously mentioned, teams of 4 students (per semester) were assigned to build an NMR probe that could be used in the Bruker 400 MHz spectrometer. Students wired, soldered, and modified machined parts to build a probe that can be tuned to deuterium and give useable spectra (Figure 1.6). They did this with the guidance of graduate students (in Dr. Mueller’s lab) and parts machined by the Department of Engineering’s machine shop (within the Eberly College of Science).  

As the project has progressed across semesters, students have improved upon the original design. For example, one team modified the coil so that it could be removed and exchanged to look at different nuclei. This type of innovation shows that students have learned to link instrument design and function. This may also be illustrative of a confidence to try new things without fear of breaking expensive instrumentation (perhaps because they now know that they can fix it).
Figure 1.6. Student-Built NMR Probe: Tuning and a Spectrum Produced. The probe below has been shown with its aluminum casing removed so that tuning rods and other components can be seen. Machined components were machined in the Department of Engineering’s machine shop (PSU). Details on the exact parts used to construct this probe and plans to produce a kit-version can be found in the reference provided.36

Although the bulk of funding went to the expansion of the NMR curriculum at Penn State university, the goal of this work was to assess these changes. This required a thorough understanding of the unique features of each module, and the goals of the curriculum overall. Once this context was established, we moved on to define key variables that will be used for assessment, so that measures could be designed and data collected.
Chapter 2

Evaluation: How Can We Measure Success of the Curriculum?

2.1 Previous Assessments for Chemistry Curricula

The *Journal of Chemical Education (JCE)* is arguably the most accessible resource for research in chemical education for chemists in the U.S., as it is distributed by the American Chemical Society and is provided to researchers and students by institutional licenses at most universities. Therefore, the curricular assessments contained within would be expected to be a model for many. A literature search of this journal revealed an intriguing trend. While there are many reports of improvements to a university’s chemistry curriculum, many of these are presented without any formal measurements. The majority of those that have been assessed rely on measures such as grades in the course and surveys of students’ attitudes to provide feedback. Assessments similar to the one used by Clough\(^37\) to evaluate a general education course are commonly used as informal indicators of progress to the faculty implementing the new program. However are there better measures that could be used for more formal published assessments?

Chemical educators have worked to construct more reliable assessments of curricula and individual lessons, but the focus of those published in *JCE* continues to be on student attitudes. While some focus only on students’ attitudes toward the
new material, others look at students’ attitudes toward science and scientific careers. A study by Oliver-Hoyo focused on the effect of anxiety on students attitudes towards learning science and chemistry. Bauer has criticized studies of attitude for chemistry curriculum assessments, and instead suggests using a measure of chemistry self-concept (a student’s beliefs regarding his/her competence) that was developed through his research.

Using student attitudes and grades on course assignments as measures may be informative for informal studies, but there are limitations to using only these to evaluate curricula for funding reports and publications. Student attitudes are not always consistent or in some cases even relevant, as seasoned teaching faculty are familiar with the many confounding factors that may make a student like or dislike a course. The same student may like one course because it is challenging, but dislike another because it is too difficult. Their grade in the course, mood during testing, and feelings towards the time of day that the class occurs are only a few of the factors that may affect ratings. Grades are constantly being monitored and affected by both students and teachers. Whether it is fitting grades to a curve directly or designing future assignments’ difficulties to aim for an acceptable mean, the social and professional implications of having too many high or low grades for a given class can make them an unreliable source of information for the educational researcher. For the NMR curriculum assessment presented in this thesis, we sought to provide a convincing argument for why these modules are an improvement using more information than grades and attitudes can provide, and we sought to do this in a logical, systematic manner familiar to chemists and scientists.
Hypothesis-driven assessment (defined in section 1.1.1) aims to use the scientific method to understand educational problems, which is the goal of chemistry education as a field. Development and implementation of an NMR curriculum at PSU was funded based on the assumption that the changes implemented would have a predictable outcome, to improve the students learning and/or overall course experience in a measurable way. Thus we focused on possible connections between course changes and student learning outcomes. Variables were chosen and manipulated to provide an experimental design that could be used to support or disprove the hypothesis. Unlike an experiment run in the lab, research within a classroom cannot be strictly controlled. Every day, students are balancing multiple courses as well as their social responsibilities, and an experiment cannot control for all of these outside effects. Even if coursework could be designed to restrict outside influences on the students, it would become useless to educators in the “real world,” who would be unable to translate these strategies to meet needs at their own universities. Thus control must be balanced with authenticity.

Also, researchers’ concerns must be balanced against classroom concerns. As educational researchers, we would like for the implemented materials and assessment measures to be as isolated as possible from other activities, since these could affect results. This may conflict with the instructor’s goal of integrating activities and presenting them in a logical sequence. Educational researchers would like to have complete control over the materials themselves, however teaching faculty should be consulted to be sure that new materials fit the goals for the course and style of the lecture or laboratory. The assessment measures must be as
unobtrusive as possible so that it does not interfere with classroom activities and assignments. In addition, incentives such as extra credit must be cleared with the instructor. Keeping all of these required balances in mind, we set about defining measurable variables and hypothesis for this assessment.

2.2 Variables of Interest and Hypothesis

The first step in designing an assessment for the NMR curriculum at PSU was to choose those variables that were most vital to the program and define these constructs. Once this was accomplished, they can be translated into operational definitions that can be measured. The variables studied in this assessment are introduced in Figure 2.1 to provide a context for further discussion.
**Figure 2.1. Variables Chosen for NMR Curriculum Study.**

**Exposure to Hands-on NMR Modules**

New NMR modules added to the curriculum must be compared to an alternative. Students will be divided into an experimental group and a control group (see below).

**Experimental group** - participants who have completed the NMR module for a given course.

**Control group** - participants who have not yet completed the new module and have instead been working on other (non-NMR related) course activities.

**Depth of Understanding**

Since modules were designed to promote greater depth of understanding for NMR concepts, this variable will confirm that modules are working as the developers intended.

**Problem solving**

To justify the time taken from other course activities, project designers would like to show that greater depth in one topic (NMR) will have benefits in addition to content knowledge in that one narrow area. Problem solving is a skill that is required for all scientists, thus improving these skills would be a worthwhile goal. There is sufficient precedent for creating measures of this skill for chemistry content.

**Confidence for NMR Tasks**

Confidence is one of many learner characteristics, which may act as an intermediary between NMR teaching and student outcomes. There are many types of confidence measures available for use/reference within the chemical education literature.
Three important factors governed the choice of variables measured as part of the study: relevance to project goals, likelihood of a measurable effect, and possibility of outside influences directly affecting the variable. The course curriculum developed must be compared to some alternative, but it would be impractical to study every addition to the chemistry curriculum brought about through this grant (as it covers three courses and in some, NMR components are spread over the full semester). Thus modules were chosen that contained the key factors introduced through the NMR curriculum. Modules assessed were hands-on, made connections to various aspects of NMR theory and instrumentation, and discussed NMR concepts in greater depth than was used in years prior to the new curriculum implementation.

A measurable effect is one that not only shows a significant difference between experimental and control groups, but also has a measurable effect size.\(^2\) Effects will be measured using statistics comparing the experimental and control groups, thus it is important that the sample size is large enough to see significant differences between groups, if they exist. Thus the special projects for Chem 426 (Chemical Spectroscopy) were excluded from this assessment because only 3-5 students completed the probe-building assignment in a given semester. Though courses earlier in the curriculum contain greater numbers of students, these students’ limited level of understanding could hinder the study. We wanted to focus on learners who are in the transitional phase between novice and expert because it is in these students that we expected to see the most change (effect sizes). Those with no prior knowledge of NMR are unable make connections and generally retain
less new information. Conversely, as learners approach the expert level, they develop organizational schemes to relate knowledge and begin to recognize patterns implicit in the area of focus. Depth of understanding requires these types of connections. Students in upper level courses are not only on their way to becoming experts in their major (chemistry, chemical engineering, etc.), but have also had multiple experiences with NMR. This allows them to build upon their fundamental knowledge and to use it in a variety of applications.

When designing experimental set-ups, the researcher must consider contributions to students’ knowledge and motivation within and outside of the course being studied. Upper-level laboratories tend to have overlapping assignments, thus modules that were more isolated and presented earlier in the course were preferred to minimize outside factors. For example, in a course such as Chem 431 (Advanced Organic and Inorganic Synthesis), students broaden their skill base with each assignment completed. This can affect their scientific understanding and confidence. Therefore the NMR module tested for this course had to be at the beginning of the semester and was in fact, their first major lab assignment due. The modules tested were tightly designed so that the majority of NMR content came from the researchers or through the assignment itself, rather than relying on outside resources (texts, TAs, instructor time). Tutorials were the primary style used for hands-on modules. However, for each course, researchers offered a brief re-introduction to NMR topics (that was not hands-on), and these were generally given in written or lecture form. For example, students in Chem 457 received a lecture where content was written and delivered by a member of the head PI’s (solid-state
NMR) research group.

2.2.1 Exposure to Hands-on NMR Modules

In the past chemistry curriculum at Penn State, students were given only limited access to the NMR spectrometer in laboratory settings, and this was a deficiency that the new curriculum sought to correct. An NMR works using physical processes that can be directed through manipulation of electronics through radiofrequency pulses. The fundamentals are often explained in textbooks through example spectra and diagrams, but rarely do students get to touch the instrument and manipulate experimental or processing parameters. In fact, prior to this project, student contact with NMR spectrometers was consistently limited to either a list of commands (80 MHz spectrometers) or auto-run modes (400 MHz spectrometer). Students were not allowed to deviate from the basic set-up, even though it is common to change parameters (and even NMR experiment types) in a laboratory setting. These restrictions are understandable considering the complexity and expense of an NMR spectrometer, and this situation is common at universities across the nation. However, this barrier may reinforce students’ fears of using the instrument, and may also force knowledge about the analytical technique to remain abstract in the students’ minds. For this reason, curriculum developers sought to introduce hands-on experiments that allowed more student contact with the NMR spectrometer and data, but that were also “safe” for the instrument itself. For example, in Chem 457, students accessed only the variables necessary for running the T₁ experiment (pulse lengths and delays) and were allowed to change only
parameters where precise values were not necessary (i.e. variable delay $\tau$ in the $T_1$ pulse sequence). In addition, the tutorial described the function of all commands and parameters entered in order to encourage students to think through interactions between the NMR spectrometer and the sample for a given process. Changing undergraduates’ perceptions of NMR was a key ambition of this new curriculum, and introduction of well-planned hands-on modules was expected to be a primary contributor to this change. This is why the assessment focused on this aspect of the new NMR curriculum.

Exposure to hands-on NMR modules is an independent variable that was controlled by the researcher. This must be compared with an alternative and so participants (in each course) were randomly assigned to either the experimental or the control group. The experimental group was made up of students who have completed the hands-on NMR module for their course. The control group consisted of students who had not yet completed the hands-on module at the time of testing (posttest), and instead had been working on non-NMR based experiment(s). Details of the experimental design used can be found in Chapter 3 (section 3.1.2).

Researchers proposed that multiple exposures to NMR content would be a contributing factor increasing students’ depth of understanding, and perhaps also their problem solving and confidence for NMR tasks. While we cannot manipulate exposure to NMR concepts outside of the courses being studied, it is important to know students’ level of exposure. Thus, this information was collected as part of the assessment. In addition, it was assured that each student had at least two exposures to NMR content by presenting materials in the course itself. The initial exposure
was given in a different format (lecture or written text), and hands-on treatment of the material was reserved for the modules described in chapter 1. For more detailed information on how this variable was treated for the assessment, see section 2.3.1.

### 2.2.2 Depth of Understanding

Though modules were specifically designed to increase depth of student’s understanding of NMR, whether or not this is occurring must be determined. Only after confirming that the new NMR modules were in fact promoting greater depth of understanding of NMR could researchers investigate how this new knowledge might be related to other learning skills and constructs. Depth of understanding was primarily used by curriculum designers to determine if they had implemented the program as intended, and so it was important to design measures that utilized their definition of the construct.

**Summary of Relevant Expertise Literature**

Through the process of learning, novices with little to no knowledge in a domain (area of knowledge) gain experience and understanding until they are at the expert level. In this study, we looked to experts in NMR to gauge depth of understanding of this content area and aimed to increase the expertise of students through the hands-on NMR modules developed at PSU. For this reason, a review of the relevant educational literature was important for the assessment. Educational researchers have devoted considerable effort to differentiating the cognitive and metacognitive processes of experts and novices. For example, Chi and Glaser\textsuperscript{42}
defined experts as having more “conceptual chunks” of knowledge, more ways of defining and relating these chunks, more connections among chunks, and more efficient ways of retrieving chunks of relevant information.

In Bransford’s book, *How People Learn: Brain, Mind, Experience, and School*, literature on expertise has been summarized and several themes emerged resulting in a model that can be used to differentiate experts from novices. Most notably, experts were found to have cognitive functions that were different from novices, specifically differing modes of organizing, recalling, and applying knowledge within their field of expertise. A cognitive process is any intellectual action that alters information within the brain and moves it from one memory store (sensory, working, or long-term memory) to another. In addition, when working within the specific domain where their expertise lies, experts’ metacognitive processes (such as monitoring and evaluating performance) were found to be more advanced than novices’ processes. Note that these differences are present only within the expert’s field of expertise, and expertise in one area does not imply the same level of expertise in any other domain.

While it is true that experts tend to have more knowledge in a given domain than novices, the way they organize this information is of greater benefit to the experts. When considering concepts or problems, experts organize ideas around “big ideas” rather than surface details. In a study conducted by Chi and Glaser, physics experts and novices were asked to sort specific physics problems into groups. In this study, it was noted that novices tended to sort problems by their surface similarities (e.g. problems with inclined planes and boxes), while experts tended to
group problems by the scientific principles (e.g. conservation of energy) needed to solve them. Rather than grouping new information in a serial manner, experts construct webs of knowledge, or **schema**, with multiple links among concepts and types of knowledge. In a study of scientific inquiry, Crismond depicts experts as having strong bi-directional links among conceptual, operational, and procedural knowledge, that are weak or non-existent for novices in that particular area of problem solving (design-redesign projects). Through the process of learning, students are at various phases of expertise, and are constructing their own schema in the domain being taught. For this reason, **prior knowledge** is of key importance to new understanding. Students who already have a more intricate web of knowledge will likely find it easier to learn new related information, whereas novices do not yet have a schema to which new information can be linked.

Experts have more information in their brains related to their area of expertise, and yet their ability to recall this information with ease and speed suggests cognitive processes different from those found in novices. Experts are able to recall greater amounts of information due to a process called **chunking**, where smaller pieces of information are “combined into larger, more meaningful chunks.” A simple example of chunking is found when trying to memorize numbers. It is easier to remember a string of ten digits if they are grouped as they are for a phone number (e.g. 0135559302 vs. 013-555-9302). In a similar manner, a person can remember a long string of letters if they are grouped into words that make up a meaningful sentence. Expertise adds meaning to otherwise random information. For example, in a study of recall and expertise, children who had played chess were able
to remember more chess pieces than college students with no chess experience, even though the college students were better at memorizing long strings of numbers. Experts are also known to have “fluent” (effortless) and automatic retrieval of relevant knowledge in their area of expertise. Since their working memory is not tied up trying to gather and relate small pieces of information, a larger percentage of the expert’s working memory can be devoted to more complex processes involved for a given task, such as problem solving.

Experts and novices also differ in their abilities to apply knowledge within a domain of knowledge. While “novices can know a principle, or a rule, or a specialized vocabulary without knowing the conditions of effective application,” an expert’s knowledge is “contextualized”. This means that the knowledge accessed by the expert is linked to the specific context in which it is relevant and useful. There is also evidence to suggest that experts recognize patterns (often within problems) which they use as cues to determine what specific conceptual or procedural knowledge is required to enable understanding (or to solve the problem). If the experts’ knowledge of cues and procedural knowledge is classified and their methods of solving particular problem type can be understood, then we can begin teaching and scaffolding these methods to improve the problem solving skills of novices. It has been found that novices tend to notice only the surface details of problems, and for this reason they will often begin working a problem that the expert can immediately identify as impossible.

Metacognitive processes have been found to be notably distinct for experts versus novices in a domain. Metacognition means that a learner is conscious of
and is exercising control over his or her cognitive processes. This control may be over basic functions such as attention or over more complicated processes such as monitoring or evaluating progress on a problem. When compared to novices in a field, experts are more likely to “rapidly check their work, accurately judge its difficulty, apportion their time, assess their progress, and predict the outcomes of their activities.” Not only are experts more adept at metacognitive processes, they are also more likely to engage in them. In a recent study of scientific inquiry and expertise, Hmelo-Silver, Nagarajan, and Day found that while monitoring statements were approximately equal for experts and novices, the experts were more likely to verbally evaluate their progress than were novices. If novices were to use these metacognitive processes more frequently, then more efficient learning and successful problem solving would likely be the results.

**Depth of Understanding in This Assessment**

In order to determine if modules developed for the NMR curriculum at Penn State were promoting depth of understanding of NMR, we designed items to measure students’ understanding of the multiple aspects of this analytical technique (applications, theory, and instrumentation). Using a 1D spectrum to determine a molecular structure is the most basic application of NMR, and researchers were concerned that this is all that the novice fully grasps. However, the expert uses in-depth knowledge to understand how more involved experiments can lead to multiple types of information about a chemical system. In turn, if students expand their view of an analytical instrument’s capabilities, then they can begin making connections
among the different perspectives (applications, theory, and instrumentation) to understand how these experiments work. Since the ability to make connections is dependent on prior exposure to NMR topics, we ensured that all participants in the study received a review of relevant NMR topics.

We expected depth of understanding to be a variable that is dependent on the experimental condition, and for those exposed to the hands-on NMR modules to have higher levels of depth of understanding of NMR. This was determined by comparing depth scores from the experimental and control groups using independent $t$-tests. In addition, we expected depth of understanding to be positively correlated to problem solving skills and confidence for NMR tasks. Correlation statistics (Pearson’s $r$) were used to find significant correlations and effect sizes. Depth of understanding presupposes a basic understanding of the topic, therefore a measure of basic NMR knowledge was also required. A positive correlation of the scores from basic knowledge and depth of understanding was expected.

### 2.2.3 Problem Solving

Curriculum designers believed that depth of understanding of NMR would have benefits beyond just additional content knowledge in that area. Evidence of this would help justify the replacement of previous laboratory experiments (at Penn State or other universities) with newly developed NMR modules. However, in what areas of undergraduates’ skills or learning could we expect to see changes?

Problem solving is a vital skill for scientists, and while this skill is rarely directly taught in an undergraduate chemistry curriculum, it is expected of any
expert in the field. The link between depth of understanding and problem solving is highlighted by Bransford’s statement that “people who have developed expertise in particular areas are, by definition, able to think effectively about problems.” Experts’ skills in organizing, recalling, and applying their knowledge all contribute to their ability to solve problems, and these are all skills that come from through depth of understanding in a specific content area. Researchers on the project felt that depth of understanding of a specific area may help students become better at problem solving in NMR, and perhaps even outside of the specific content area. This is an extension of the premise upon which graduate research programs are based. When completing a graduate degree, it is expected that by studying one small area of chemistry (one research project) in great depth, the resultant “doctor” will be an individual fully capable of in-depth problem solving for a range of situations in their future career. Supporting this claim, Ph. D. chemists are often hired and are very successful in careers where their daily activities are completely unrelated to their graduate research. Relating depth of understanding to problem solving skills in an undergraduate setting would give evidence to the benefit of this NMR curriculum, and may encourage faculty at other universities to adopt this type of program.

Since problem solving is an essential skill for any scientist, it is not surprising that many instruments have been developed to measure this construct. In fact, literature precedent for measuring problem solving in chemistry courses was another reason that this variable was chosen for study. As will be discussed later (section 2.3.3), a student’s ability to solve problems is limited by transfer, and so an instrument was developed to measure the construct within the context of NMR. One
indicator of expertise in a specific content area is the expert’s ability to use the knowledge that they have to solve problems; therefore we will be looking for evidence of these developing skills. Details on how this variable was measured can be found in section 2.3.3.

In this study, problem solving skill is a dependent variable. Multiple factors may contribute to this skill, and those considered most relevant to this assessment were exposure to NMR modules, depth of understanding, and confidence for NMR tasks. Thus, we were looking for differences in the problem solving scores for the experimental and control groups, and significant correlations between this variable and both depth of understanding and confidence for NMR tasks.

2.2.4 Confidence for NMR Tasks

The study of learner characteristics arose from the observation that not all people gather and process information in the same way. There are many factors which can affect a given learning experience, and these may be either stable or changing. Stable characteristics are attributes such as gender, cultural factors, and I.Q, which remain relatively unchanged over time. Other learner characteristics change frequently or with regard to area of study. These are factors such as prior knowledge, attitude toward a topic, and motivation. By altering certain learner characteristics, a student’s performance on an academic task may be improved.

Analytical instruments, especially complex ones such as NMR spectrometers, can be intimidating to undergraduates, and this apprehension must be overcome for a student move towards the expert-level in this technique. Confidence is a learner
characteristic that may be key to NMR learning. Specifically, this study sought to measure confidence for NMR related tasks, since confidence is often task specific, rather than an all-encompassing trait for a given discipline. Exposure to hands-on modules and increased understanding of NMR topics could both be expected to increase student confidence, and this construct may be an intermediary between the modules and student performance in a problem solving task.

Confidence for NMR tasks is another dependent variable. We believed that students who completed hands-on NMR modules would be more confident in NMR tasks than those who have not. Correlations to depth of understanding and problem solving skills were also implied predicted by the hypothesis of this work. In addition, it was proposed that these modules might also have an effect on students' confidence for general science tasks. Therefore, statistical analyses similar to those used for previously described variables were also performed for confidence for NMR tasks and confidence for general science tasks scores.

2.2.5 Hypothesis and Research Questions

The hypothesis for this project is that deeper understanding of one particular analytical technique (NMR) will increase undergraduate students’ abilities to solve chemical problems. This is based on the assumption that use of this curriculum leads to increased depth of understanding (of NMR topics). Confidence for specific tasks may be an intermediary between changes in depth of understanding and problem solving skills, and whether confidence is linked to these two variables is a sub-question of this research. The hypothesized relationship among the chosen
variables is shown in Figure 2.2.

Figure 2.2. Hypothesis Used to Assess NMR Curriculum.

In order to determine if depth of understanding is related to problem solving skills in the participants, instruments were developed to measure depth of understanding of NMR and problem solving within this context. Since confidence for NMR tasks is a learner characteristic that may arise through understanding and may positively affect problem solving, this construct was also measured. Testing of this hypothesis was dependent on the prior development of hand-on NMR modules that promote deeper understanding, the product of the CCLI grant received from the NSF.

This hypothesis requires more than one research question to be answered, and several sub-questions will also need to be addressed. The research questions that will be addressed in this assessment are:
1. Primary research question: Is depth of understanding of NMR linked to problem solving skills?

2. Are the NMR modules working as intended? Do they promote depth of understanding of NMR?
   a. Will students who complete the NMR modules have a greater depth of understanding of NMR than students who do not complete the modules?
   b. Is depth of understanding increasing over the course of the experiment?

3. Is confidence an intermediary between depth of understanding and problem solving skills? Is it linked to both variables?

4. What levels of confidence are affected by the NMR modules?
   a. Will confidence for the NMR class skills used in the modules themselves be greater for those who have completed the modules?
   b. Will confidence for NMR tasks not practiced in the course be affected?
   c. Will confidence for general science tasks be affected?
   d. Are different levels of confidence (class skills, NMR tasks, general science tasks) linked to each other?
2.3 Measuring the Constructs: Instruments Used and Developed

Since many of the outcomes of this curriculum are not tangible, it was necessary to design instruments to quantify any changes in the variables. When creating an assessment with multiple variables, it is crucial to be clear on what program designers meant by terms used in their goal statements for the project. A constitutive definition defines the term in relationship to other terms, and is often similar to what would be seen in a dictionary. However, where these were too broad, discussions with the grant-writers lead to more specific definitions. In addition, by asking these researchers to give examples allowed meaning to be placed within the context of NMR, and this was often used in place of a more broad definition. For example, by asking, “what does a student with deep understanding of NMR know that is different from one without?,” we were able to clarify what the grant-writers meant by depth of understanding of NMR. An operational definition consists of “actions, processes, or operations used to measure or identify” examples of the term being defined. Observable events, rating scales, or scores on instruments could each serve to provide an operational definition. Both constitutive and operational definitions of the variables used in this study can be found in Figure 2.3.
Figure 2.3. Variables as Defined by Researchers and Assessment.
When designing instruments and methods to measure the chosen variable, it was crucial use the terms as defined by the developers of the NMR curriculum. These definitions are shown below. Simply stated, constitutional definitions are what the construct means, and operational definitions are what will be measured for that variable.

<table>
<thead>
<tr>
<th>Variable/Term</th>
<th>Constitutive Definition</th>
<th>Operational Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on NMR</td>
<td>Time spent manipulating input or output for an NMR experiment</td>
<td>Whether or not students have completed the new NMR module in their course</td>
</tr>
</tbody>
</table>
| Depth of Understanding NMR                | 1. Making connections within and among the various aspects of NMR (applications, theory, and instrumentation)
2. Knowing types of information about a chemical system that NMR can give | Score on depth of understanding items (free-response)                                   |
| Problem Solving                           | Approaching a complex problem in a logical way                                          | Score on HETCOR problem (free-response)                                                |
| Task Confidence                           | Belief in one’s capabilities to successfully complete a task, specifically for NMR tasks | Self-reported scores on a 7-point Likert scale                                         |

2.3.1 Exposure to Hands-on NMR Modules

Exposure to hands-on NMR modules was a manipulated variable, rather than one to be measured. Details of the experimental design used to manipulate this variable can be found in chapter 3 (section 3.1.2). We controlled this variable through the materials created and the layout of the particular course being tested. The researchers defined hands-on contact as time spent manipulating input or output for an NMR experiment. When manipulating input, undergraduates should
be learning what the instrument can do and how changing parameters can alter results. Students can learn in detail what information is available from an NMR spectrum by processing their data and modifying the output to most clearly depict important information. The modules chosen for testing were the processing tutorial and “Analysis of a Crude Product” assignment in Chem 431 and the T₁ experiment in Chem 457. These were previously described in Chapter 1 (sections 1.4.1 and 1.4.2). The two activities in Chem 431 focused on output, while the T₁ experiment for Chem 457 focused more heavily on inputs for the NMR experiment. The operational definition of this variable was based on completion of the new hand-on modules: students who had completed the modules prior to testing (posttest) constituted the experimental group, and those who had not made up the control group.

Before testing began, we met with teaching faculty to discuss their needs and where our modules and testing could fit to the course schedule with minimal overlap of content. For Chem 457, the T₁ experiment was designed to fit within one lab period. An average student spent approximately 45 minutes setting-up and running the experiment, and 30 minutes processing the data. There was a limit to the number of students that could perform the lab at one time, because we wanted each student to have direct contact with the instrument. Thus, students completed the lab in pairs, two pairs per week of lab. It was customary for different groups of students to complete experiments at different times in Chem 457, and this assignment fit well into the laboratory rotation. There were no other NMR related experiments or materials in this course, outside of the content presented as part of our study. For Chem 431, the processing tutorial was completed outside of the
normal lab period and took 30-40 minutes to complete. However, the “Analysis of a Crude Product” assignment required work during several lab periods, and students worked at their own pace for this assignment as was usual for this advanced course. The estimated time spent processing the data from this Crude Product assignment is 30 minutes. Analysis of these data to determine the components and ratios of the mixture might take considerably longer, depending on the difficulty of the unknown and the student’s understanding of NMR. Placing the module to be assessed at the beginning of the semester meant that the only overlapping lab activities were other forms of training, a quiz, and possibly work on a course-long project. These assignments were not NMR related, so interference was minimized. It should be noted that this course does depend heavily on NMR, and so testing (posttests) had to take place before the majority of course assignments, since most required the use of NMR in some capacity.

The experimental and control groups differed in their exposure to hands-on NMR modules, but this was not the only feature of the new curricular plan that had to be considered. Repeated exposure to NMR concepts was deemed necessary to allow for testing of depth of understanding and to ensure that students from different levels of prior exposure (previous courses that contained NMR) were brought to a more even level. For this reason, all students were given some NMR teaching prior to testing. This initial exposure was not hands-on, but consisted of a lecture (Chem 457) or written summaries (Chem 431), so we were still testing the effects of hands-on learning since the control group has had none.

As previously stated, depth of understanding requires prior knowledge in the
content area for learners so that they can relate this to new concepts being learned. Not all students have had the same level of exposure to NMR prior to the courses being studied, so we collected information on students’ previous exposure to NMR content. This was obtained by asking students which courses they had taken that contained NMR content, and what NMR techniques (experiments) they had learned. The items used for the final pretest assessment can be found in Appendix C (section 1).

In this study, most students have had at least one course that covered some NMR concepts in the past (due to the prerequisites for these upper level courses), but effort in previous courses and time in between courses can affect retention. Requiring more than one exposure to NMR concepts within the course ensured that all students had an adequate base of NMR knowledge. The initial exposure in each course was a brief re-introduction to NMR topics and provided a context for the information to be learned as part of the hands-on modules. These introductions were given in written or lecture form. Students in Chem 457 received a lecture that covered some quantum theory and modern applications of NMR experiments. In Chem 431W, students were given an overview of the NMR spectrometer, a demonstration about making and running NMR samples (in automation), and a quiz requiring basic NMR spectra interpretation skills covered in prerequisite courses, Chem 36 and 39. Control groups received only this initial exposure and no hands-on experience. This was designed to test the key aspects of this new curriculum, in-depth and hands-on treatment of NMR by ensuring that the experimental group had further exposure to NMR in multiple contexts and hands-on experience that was not
offered to the control group (until after testing was complete).

2.3.2 Depth of Understanding and Basic NMR Knowledge

Though it is evident that experts in the field have an in-depth understanding of NMR, this encompasses a broad range of knowledge and skills, and these had to be narrowed down before measurement could begin. This assessment sought to focus on those areas of understanding that the grant-writers found lacking from previous chemistry curricula. One of the primary goals written into the grant\textsuperscript{5} for this project was to build upon and link the multiple aspects of nuclear magnetic resonance (applications, theory, and instrumentation). Students with a greater depth of understanding should be more capable of making connections within and among the three principal areas listed. Authors\textsuperscript{5} also highlighted the need for students to understand more than the most basic applications. According to NMR experts, novices tend to over-simplify analytical instruments by narrowing their use down to the most basic function. In the case of NMR, this would be to determine the structure of a molecule through a 1D (H\textsuperscript{−}NMR) spectrum. An expert can relate how the instrument works to the multiple functions that it can perform. Relating this to the variable at hand, students with a deeper understanding should have a broader view of types of information about a chemical system available through NMR experiments.
**Depth of Understanding Items**

When considering how best to measure depth of understanding of NMR concepts, we decided to use **free response items**. For these items, students are given an open-ended question or prompt. These can lead to multiple answers, which can give more information about what students understand. We wanted to know how students thought of NMR: not just what they knew, but also what they deemed relevant and how they grouped information. By using open-ended questions, we intended to keep from leading students into constricted answers.

Two free-response items were chosen to measure depth of understanding (of NMR), in order to cover both aspects of the construct (shown in Figure 2.4). Determining whether or not students are able to connect information regarding the different aspects of NMR (applications, theory, and instrumentation) was a difficult prospect. Yet, it was clear that the focus of these aspects was an overall picture of the analytical technique as a whole. This lead to the first question, “In general, what is NMR? What does it do?”, which will be referred to as **depth Q1**. Since a sound understanding of NMR as a whole requires the synthesis of different aspects, this was a compound question. The first half of the question, “What is NMR?”, was expected to lead to answers such as what the acronym stands for (Nuclear Magnetic Resonance), parts of the spectrometer, and/or how the magnet interacts with the sample molecule. The second half, “What does it do?”, could lead to answers such as the purpose of the instrument to find chemical information or a description of the processes that occur during an experiment. Determining how much students understood about an NMR spectrometer’s applications was a more straightforward
question, which the second item, “What types of information about a molecule/reaction can you get using the NMR spectrometer?” (depth Q2), aimed to answer. This question was designed to force students to think beyond a list of experiments (acronyms, such as COSY, NOESY, etc.) and really think about what these experiments can tell a chemist.

**Figure 2.4. Depth of Understanding Items.**

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
<th>Designed to Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth Q1</td>
<td>In general, what is NMR? What does it do?</td>
<td>Student connections within and among the various aspects of NMR (applications, theory, and instrumentation)</td>
</tr>
<tr>
<td>Depth Q2</td>
<td>There are many types of NMR experiments. What types of information about a molecule/reaction can you get using the NMR spectrometer? Please LIST as many as you can think of. (e.g. the structure of a molecule, etc.) Do NOT give details about the experiment!</td>
<td>Student knowledge regarding the types of information about a chemical system that NMR can give</td>
</tr>
</tbody>
</table>

**Basic NMR Knowledge Items**

For any content area, prior knowledge is expected to be a substantial contributor to new understanding. As a person gains more information about a particular subject, they can begin organizing this information into logical patterns or webs of knowledge relating concepts. Even if all participants in the study have had prior exposure to NMR content in previous courses, however exposure does not ensure that the material was learned or retained. For this reason, a series of items were created that covered the fundamental concepts of NMR that students should
have already encountered. Experts in the field of NMR and teaching faculty for participating chemistry courses were consulted to assure that items covered concepts that were both important within the area of NMR and relevant to what was being taught.

The content and breakdown of items included for the basic NMR knowledge section of the assessment are shown in Figure 2.5. For a complete listing of items used in the final assessment, see Appendix B (section 1, parts I and II). This information was taught in organic chemistry lecture and lab courses that were prerequisites for Chem 431. All chemistry majors in Chem 457 have taken these organic courses, and a survey of participants showed that almost all had taken Chem 431 prior to the study. Basic NMR knowledge items were used to measure two types of knowledge, the ability to interpret a $^1$H NMR spectrum (reading spectra) and the ability to correctly define concepts that are foundational to NMR knowledge (definitions). There are multiple types of information available from a 1D spectrum, and students should be able to interpret these as well as relate them to the molecule that gave rise to the NMR signals. For items in the reading spectra category, students were given examples of 1D $^1$H NMR spectra and asked questions about them, either in multiple choice or short answer format. Some items were weighted (worth 2 points instead of 1 point), and all were scored as either correct (maximum point value) or incorrect (0 points). Items in the definitions category required students to relate general features of an NMR spectrum to features of molecule from which the peaks arise. Definition items used the following format,

In general, for a $^1$H NMR spectrum:
What does a peak's _____ tell you about the corresponding hydrogens?

where an NMR term (shift, splitting, integral, or coupling constant) was present in the blank. Responses to these short answer items were expected to range from partial understanding to complete understanding, and partial credit (1 out of 2 points) was possible.

**Figure 2.5. Item Summary for Basic NMR Knowledge Section.**

These items tested students’ abilities to read a $^1$H NMR spectrum and to define concepts that are foundational to an understanding NMR. This information was taught in courses that are prerequisites for both Chem 431 and Chem 457, but this measure determines what information was retained.

<table>
<thead>
<tr>
<th>Item Category</th>
<th>Number of Items</th>
<th>Total points (Item Numbers, Point Values)</th>
<th>Type(s) of Items (Item numbers)</th>
<th>Concepts Covered (Number of Items)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading Spectra</strong></td>
<td>9</td>
<td>12 (1-6, 1 pt each; 7-9, 2 pts each)</td>
<td>Multiple choice (1-7)</td>
<td>Phasing (2) Shifts (2) Splittings (2) Integrals (2) Types of H’s (1)</td>
</tr>
<tr>
<td><strong>Definitions</strong></td>
<td>4</td>
<td>8 (1-4, 2 pts each)</td>
<td>Short answer (1-4)</td>
<td>Shifts (1) Splittings (1) Integrals (1) Coupling Constants (1)</td>
</tr>
</tbody>
</table>

**2.3.3 Problem Solving**

Many factors must be considered when designing a problem solving task. When solving a problem, students are often limited by their abilities to transfer prior knowledge. Even though the problem strategies may be the same, transfer will be more difficult for tasks that the students consider “too different” from what they have encountered in the past. This perceived difference in a new task and prior
tasks is called transfer distance. For this reason, it is often necessary to design a problem solving task that is close in transfer distance to the information being learned as part of the curriculum being assessed, rather than using an existing measure. Thus, we designed a task that was similar to the hands-on NMR modules for both courses assessed: the XWIN tutorial for Chem 431 and the T₁ experiment for Chem 457. The ideal case was to use one task to measure problem solving for both courses.

The goal was to have a task that was similar, but not too similar to what the students had already done. This is because tasks which are very similar to previously learned material and that can be solved through an algorithm of “well known and practiced procedures” are deemed exercises, not problems. Problems require a higher level of thinking. For example, typically math texts provide exercises at the end of a chapter. These are designed to allow students to practice using the equations and methods taught throughout the chapter. Even difficult exercises can be solved using steps taught in the current chapter and perhaps previous chapters. In contrast, problems in the same area would ask students to use these same equations in a different context, and in a way that requires other types of thinking. This could be illustrated by a word problem that uses a real-life situation to ask a question that is solved using the math equations being studied. Here students must use not only their newly acquired math skills, but also their verbal skills (abilities to interpret what information the word problem is asking for and giving them). Once they know what the problem is asking, they must then translate the written information into the correct format needed to use these math
skills directly. However, context is important and if students have already learned this process of interpreting word problems through the course being studied (or prior to this), then this may not be considered a problem. This is because educational researchers have added the requirement of a higher cognitive load to the distinction of a problem versus an example. Therefore, what may be a problem (high-order cognitive load) for one student may be an exercise (low-order cognitive load) for another. The HETCOR problem used to measure problem solving for this study fits both of the aforementioned criteria (for both courses).

Heteronuclear Correlation (HETCOR) is an NMR experiment that reveals which hydrogens are directly attached to each carbon (or other heteronuclei). This is possible due to through-bond scalar couplings (e.g. $J_{X-H}$) present between any two NMR-active nuclei that are part of the same bonding network. For this experiment, the spectrometer is set-up to detect heteronuclear couplings (e.g. $^1$H-$^{13}$C or $^1$H-$^{15}$N), rather than homonuclear couplings (H-H). The focus of the HETCOR problem used for the problem solving measure was the output. For this experiment, a 2D plot is generated where the x and y axes show the 1D spectra of each of the relevant nuclei (in this case $^1$H and $^{13}$C) and the HETCOR spectrum itself consists of points where the corresponding shifts are coupled (meaning the $^1$H and $^{13}$C atoms are attached). For example, in Figure 2.6 the hydrogen represented by the quartet at 4.2 ppm on the $^1$H-NMR spectrum is connected to the carbon represented by the peak at 60 ppm on the $^{13}$C-NMR spectrum. HETCOR spectra not only allow students to compare a $^1$H and $^{13}$C spectrum in close proximity, but also to use the information in the 2D portion of the spectrum to access additional information.
**Figure 2.6. HETCOR Spectrum Used in Assessment.** This is a HETCOR spectrum of the molecule used for the problem solving task in this assessment. Note that the $^{13}$C spectrum is on the $y$-axis, and the $^1$H spectrum is on the $x$-axis. Above the 1D $^1$H spectrum is shown the multiplicity (splitting) and integral (relative number of hydrogens) for each peak for this spectrum. All peaks in the $^{13}$C spectrum are singlets (1H) as is usual for these types of spectra.

Key
- d = doublet, t = triplet, q = quartet, m = multiplet; dd = doublet of doublets
- (#H) = integrated value, relative number of hydrogens.

We chose the HETCOR problem because connections can be made between this task and the NMR modules for both courses being studied. By using the same problem solving task, we can compare the results between courses. We expected...
students to focus on different aspects of this problem depending on which course they were currently taking, but answers were not restricted in any way. We expected students in Chem 431 (Advanced Organic Chemistry Lab) to be able to interpret the 1D \(^1\)H and \(^{13}\)C spectra along the axes of the HETCOR spectrum, and to be less likely to focus on the 2D area. We expected that Chem 457 (Physical Chemistry Lab) students would understand reading a 2D layout for results (similar to the T\(_1\) experiment), but would be less likely to focus on structural information.

It was also important that the problem solving task be novel for all participants. The HETCOR spectrum can only be a problem for those students who have not been formally trained to interpret it. While HETCOR experiments are not part of the curriculum for any of the courses under study, there is an elective course called Structural Analysis of Organic Compounds (Chem 439) that some participants may have taken. There is no hands-on experience in this course, however NMR spectra of various types (1D and 2D) are interpreted and used within the context of structural elucidation. Therefore, we chose an experiment that had not been covered within this course.

The problem solving task was presented through a free response item. We consulted experts (both experts in NMR and the course instructors) to confirm that the difficulty level of this specific problem was appropriate for the student sample that would be participating in the study. For the complete problem solving task as it was presented in the final posttest assessment, see Appendix B (section 3). Participants were given instructions for the problem solving task and then asked to list “all the information” about the molecule that they could gather from this
spectrum (item 1). Following the initial request for information, students were asked to think about the similarities and differences among this experimental output and the activities completed in their NMR module (items 2 and 3 of this section). These were not scored, but were an attempt to help students make connections among learned content. After these guiding questions, students were asked to supply any new information they had determined (item 4), presumably after making connections to their previous NMR experiences with the hands-on modules.

One limitation of online testing for this section was that students could not draw organic structures, but instead had to use text to convey their knowledge. This is atypical for spectral assignment problems used in coursework (either lecture or lab courses). In most cases, students instead use drawings to convey the final structure or pieces of the carbon framework. Students often do not learn the language to discuss the structural features for this reason, and so this was provided to them. A visual dictionary accompanied this problem, which linked structural terms to their drawn equivalent (see Appendix B).

2.3.4 Confidence

Confidence is a broad term that needs to be narrowed for this work, since in the education research community, there are several related concepts: self-concept, self-efficacy, and task confidence. The most general of these concepts is self-concept, defined as a “cognitive appraisal of physical, social, and academic competence.” While a student’s beliefs about their physical or social competence might have an effect on their self-esteem (an emotional reaction to the student’s
view of his or herself), academic self-concept is the area that most directly affects student achievement. Academic self-concept may be subject-specific, but it is still very general. A student that states that she is “good at math” has high self-concept in that area of study, but we still do not know how confident she feels completing a specific math task such as integrating an equation with exponents.

The most task specific type of confidence is self-efficacy. First defined by Bandura,\textsuperscript{55} self efficacy is defined as a person’s “beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives.” Self-efficacy relates to the capability to accomplish a task, rather than just a person’s innate ability. In addition to the latter, effort and persistence are also included in the learner’s judgment of whether or not they can successfully complete a task. Not only should this type of confidence relate to a task rather than a general area, but it must also be specific and fit within the context of the study. Pajares\textsuperscript{56} has warned that by trying to encompass too many aspects of an academic task it becomes vague, and a generalized set of items does not accomplish the same goals that self-efficacy items should. Self-efficacy measures should minimize the distance from the task to measurable outcomes of a study. The more closely the tasks/skills described in self-efficacy items are to the measured outcome, the more highly correlated the resulting scores will be. Here is where context plays a key role. For example, “Rate your confidence to make an A in a chemistry course” is neither specific nor relevant to the research questions for this NMR assessment project. However, it would be applicable if the study was aimed at increasing students’ ability to make an A in a chemistry course. Items for a self-efficacy measure are self-
reported ratings of the strength of the participant’s confidence that they can successfully complete a task. The items themselves should be based on skills (and subskills) that are required to accomplish the outcome being measured.57

This study required a level of confidence more task specific than self-concept in science but more generalizable than self-efficacy, and so we sought to measure something we called “task confidence”. It has been noted that while self-efficacy is quite precise, it can be of limited practical relevance.58,59 In some cases, researchers want to know if their curriculum has a broader impact and can affect confidence for more general skills or tasks. A task confidence measure contains items that relate to tasks, but here the tasks (or subskills) listed are aimed at defining those skills that an exemplary student would have in the particular area being studied. For the NMR curriculum assessment, we designed items that described the skill set of competent students of NMR, chemistry, and science. Another distinction of this type of confidence is that the tasks described are not measured as an outcome of the study. Henceforth, we will refer to these types of items as task confidence items.

In this study, all confidence items were designed to measure task confidence (Appendix B, sections 2 and 5). Items that measure confidence for NMR tasks asked at what level students feel confident completing tasks that an NMR spectroscopist would be expected to complete as part of their job. Increasing confidence for these tasks could help the students in future work, even though they were not completing these tasks as part of the course. We were also interested to know if confidence could be extended to even more distantly related tasks, and so items that measured confidence for general science tasks were also included on this assessment. Students’
self-reported confidence for NMR skills learned in the modules (confidence for class skills) are the most closely related to the study conditions. However, since we are not directly measuring whether or not the students actually have these skills, we will still consider these to be task confidence items (rather than self-efficacy items). It should be noted that we would expect the scores for participants’ confidence for class skills and their abilities to successfully complete the same skills as part of the NMR modules to be highly correlated.

Instrument development began by modifying the corresponding sections of the Chemistry Attitudes and Experiences Questionnaire (CAEQ). This questionnaire was developed by Dalgety to measure student attitudes toward chemistry, confidence related to chemistry tasks, and classroom experiences. Only items from the “self-efficacy” measure were used, though it should be noted that these were designed to be used for a study of a first-year chemistry course to determine factors that affect future enrollment in chemistry courses. While this was referred to self-efficacy for the Dalgety study, for our context these items are too general to be considered as such, and were used as task confidence items. The “self-efficacy” section of the CAEQ was shown to have a test-retest reliability of 0.95 using Cronbach’s alpha. Correlations were also calculated using Pearson’s $r$ to see if variables were linked. For reference, $r = 0$ indicates no correlation, and $r = 1$ represents perfect correlation. The “self-efficacy” measure in Dalgety’s study shows significant Pearson’s $r$ correlations ($p < 0.01$) to chemistry learning experiences: lectures ($r = 0.38$), tutorials ($r = 0.29$), practicals ($r = 0.47$), and demonstrations ($r = 0.34$). This suggests predictive validity since the theoretical basis of the study would
predict these experiences to be correlated to self-efficacy in chemistry.

Task confidence items for general science tasks and NMR tasks were developed using the CAEQ. The complete list of items used for the final assessment can be found in Appendix B (section 5), however a sampling of these items is shown in Figure 2.7. Each item was scored on a 7-point scale where 1 = not confident and 7 = totally confident. The general science tasks were taken directly from the CAEQ instrument, but only those related to either experimental/laboratory skills or the ability to learn or understand chemistry concepts were used. These CAEQ items were also used as models to design the confidence for NMR task items. It is important to be sure that the items used for an instrument cover all aspects of the construct/variable they are designed to measure (construct validity). After consulting with NMR experts and teaching faculty, the most relevant skills were chosen and were confirmed to be representative of the skills required for a competent NMR spectroscopist (Figure 2.7). It should be noted that the goals of this project were not to train every student in the technical skills required of an expert in NMR, and so only the relevant scientific abilities (e.g. designing experiments, interpreting data, making conclusions, etc.) were included for this measure.
Figure 2.7. Task Confidence Items for General Science and NMR Tasks. Sample items from the confidence sections of the NMR assessment are shown below.

Items for the confidence for general science tasks section were chosen to represent necessary skills for a competent scientist/chemist. These were taken from the “self-efficacy” section of Chemistry Attitudes and Experiences Questionaire (CAEQ). Items 1 and 2 are exemplary of items that measure students’ confidence for experimental/laboratory skills. Item 3 is representative of items that measure students’ confidence in understanding chemistry concepts.

Items for the confidence for NMR tasks section were chosen to represent necessary skills for a competent NMR spectroscopist. These items were modeled upon items from the CAEQ, but were modified to relate to NMR. Items 4 and 5 are exemplary of items that measure students’ confidence for experimental/laboratory skills. Item 6 is representative of items that measure students’ confidence in understanding NMR concepts.

<table>
<thead>
<tr>
<th>Confidence for General Science Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items were taken from the CAEQ instrument.</td>
</tr>
</tbody>
</table>

**Sample Items**

Please indicate how confident you feel:

1. Ensuring that data obtained from an experiment is accurate

2. Proposing a meaningful question that could be answered experimentally

3. Explaining something that you learned in a chemistry course to another person

<table>
<thead>
<tr>
<th>Confidence for NMR Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items were formulated by modifying items from the CAEQ. Informal interviews with NMR experts and teaching faculty confirmed that skills were representative of those required for a competent NMR spectroscopist (content validity).</td>
</tr>
</tbody>
</table>

**Sample Items**

Please indicate how confident you feel:

4. Designing an NMR experiment to answer a specific research question

5. Running an experiment on the NMR spectrometer without guidance

6. Explaining basic NMR theory to a friend not in this course
We would not expect students’ confidence for general tasks (NMR or science) to be changed through participation in this study, unless confidence for tasks performed as part of the NMR modules were concurrently affected. Teaching faculty and NMR experts worked together to come up with a comprehensive list of the skills that could be gained through each hands-on NMR module. The final set of items for confidence for class skills (NMR) can be found in Appendix B (section 2), and a sampling of these items is shown in Figure 2.8. This set of items was unique for each course (Chem 457 and Chem 431) since NMR modules were tailored to fit course content and objectives. Only items that referred to NMR skills that could be applied in future work and/or were likely to be used in or outside the course at some point in undergraduate career were used for this measure.
Figure 2.8. Task Confidence Items for Class Skills. Sample items for confidence for class skills (NMR) are shown below. These were designed to represent the skills needed for a competent student to complete the NMR module successfully. All items were designed to measure students’ confidence in for specific tasks completed within the modules.

<table>
<thead>
<tr>
<th>Confidence for Class Skills (NMR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items were designed to match those skills used in the hands-on NMR modules. Class skills depend on the course under study (Chem 457 or Chem 431). Informal interviews with NMR experts and teaching faculty confirmed that skills were representative of those required for a competent student to complete the NMR module successfully (content validity).</td>
</tr>
</tbody>
</table>

Sample Items for Chem 457
Please indicate how confident you feel:

1. Setting up an NMR experiment using X-WIN software
2. Phasing for a T1 experiment
3. Relating tau-c values to atomic motion in a molecule

Sample Items for Chem 431
Please indicate how confident you feel:

4. Phasing an NMR spectrum
5. Integrating peaks on an NMR spectrum
6. Determining ratios (relative amounts) of different compounds in a mixture using NMR alone

Instrument development is a vital part of the assessment process, and considerable time was spent constructing measures of all relevant variables. These descriptions represent the testing measures used for the final assessment (used in studies 1 and 2), but some additional items were present during the pilot study. Good testing instruments can still be used poorly and result in assessments that provide misleading information. Without well thought-out testing methods and experimental designs, good measures are wasted. Hence, we strove to ensure the
best methods of testing and scoring possible within the constraints of our assessment. These are described in the next chapter.
Chapter 3

Methods Used for Testing, Scoring and Data Analysis

3.1 Testing Methods and Experimental Design

3.1.1 Online Testing

A summary of the instruments used for this NMR curriculum assessment are shown in Figure 3.1. Variables measured on the posttest included depth of understanding, basic NMR knowledge, problem solving (HETCOR problem), confidence for NMR tasks not performed as part of the course, confidence for class skills used in the hands on NMR modules, and confidence for general science tasks. The development of these measures was described in Chapter 2 (section 2.3). The final version of tests used to assess Chem 457 and Chem 431 are shown in Appendices B (posttest) and C (pretest).
**Figure 3.1. Summary of Instruments Used for NMR Curriculum Assessment.** Measures used in the undergraduate NMR curriculum assessment at Penn State are shown below. For each measure, the number of items, range of possible scores, item type, and intent of measure is given. All measures were administered through an online survey. Measures that were given on both the pretest and the posttest are denoted **Pre/Post**. Those that were only given on the posttest are denoted **Post Only**.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Number of Items</th>
<th>Min to Max Score</th>
<th>Type(s) of Items</th>
<th>Designed to Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depth of Understanding</strong></td>
<td>2</td>
<td>0 to 11</td>
<td>Free response (open-ended)</td>
<td>Can students make connections about the various aspects of NMR (depth Q1); Do students know what kinds of information NMR can give (depth Q2)</td>
</tr>
<tr>
<td>(Pre/Post)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Basic NMR Knowledge</strong></td>
<td>13</td>
<td>0 to 20</td>
<td>Multiple choice, Short answer</td>
<td>Student’s knowledge of basic NMR concepts, such as: reading spectra, integrals, phasing, and splitting</td>
</tr>
<tr>
<td>(Pre/Post)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Problem Solving: HETCOR Problem</strong></td>
<td>1</td>
<td>0 to 5</td>
<td>Free response (open-ended)</td>
<td>Can students solve a HETCOR problem solving task (or approach the problem in a logical way)</td>
</tr>
<tr>
<td>(Post Only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confidence for NMR Tasks</strong></td>
<td>5</td>
<td>5 to 35</td>
<td>1 to 7 point rating scale</td>
<td>Student confidence for NMR tasks NOT performed as part of the coursework</td>
</tr>
<tr>
<td>(Pre/Post)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confidence for Class Skills (NMR)</strong></td>
<td>Chem 457: 9</td>
<td>9 to 63</td>
<td>1 to 7 point rating scale</td>
<td>Student confidence for NMR tasks completed as part of the hands-on NMR modules</td>
</tr>
<tr>
<td>(Post Only)</td>
<td>Chem 431:10</td>
<td>10 to 70</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confidence for General Science Tasks</strong></td>
<td>6</td>
<td>6 to 42</td>
<td>1 to 7 point rating scale</td>
<td>Student confidence for general science tasks</td>
</tr>
<tr>
<td>(Pre/Post)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This study used an experimental and a control group to compare variables, but in order to determine if students’ scores were increasing over the course of the experiment, it was necessary to compare a pretest and posttest score for each participant. In addition, the pretest can be used to confirm that the experimental and control groups being tested are essentially equivalent. The posttest assessment contained measure of all variables, but some measures were excluded from the pretest. The problem solving task (HETCOR problem) was only given on the posttest. For this problem, researchers feared testing interference, a validity threat where more than one exposure to the problem could itself lead to increased scores. Excluding the confidence for class skills measure from the pretest was done for a different reason. Researchers wanted to know if performing the NMR tasks as part of the hands-on NMR modules affected students’ confidence for these same skills. At the point of the pretest, none of the students had performed any of the tasks listed. So instead of measuring this variable before and after, we planned to compare the experimental and control groups to see if there was a difference in confidence for those who had completed the NMR modules. Since the groups were randomly assigned (see section 3.1.2), these two methods should have approximately equivalent results.

The pretest also contained items used to gather demographic and academic information about the participants (see Appendix C). Information on students’ major/concentration were collected to confirm the sample demographics expected for the courses. Students were also asked their career goals as these may contribute to their motivation for working on the NMR modules and other assignments. Since we
cannot control every aspect of a student’s knowledge of NMR, it was important to
gather information about their background knowledge regarding this content. The
primary contributor of a participant’s prior NMR knowledge was the courses they
have taken. The content of courses at Penn State is very consistent for lower-level
and prerequisite courses. Where knowledge of the exact NMR components present in
a course were lacking, interviews with faculty teaching those courses were used to
supplement this information. In addition, we asked participants which NMR
experiments they were familiar with, in case some students had exposure to NMR
outside of classes (e.g. undergraduate research). A vital aspect of the new NMR
curriculum was increasing hands-on contact with the NMR (input and output).
Therefore, we also used the pretest to determine what spectrometers the students
had used, and what training on these instruments they had received.

Surveys at Penn State were administered through an online testing site in
collaboration with the University Testing Services (UTS), which is operated by the
Schreyer Institute for Teaching Excellence.62 This Penn State center facilitates
undergraduate teaching by providing educational testing, learning assessment
instruments, and teaching effectiveness feedback for faculty at the university. We
consulted with Ralph Locklin, director of the University Testing Services, and Sarah
Zappe (research assistant at the Schreyer Institute) to determine the appropriate
format of items and uploading procedures for online tests. Once the initial survey
and class lists for the courses (Chem 457 and Chem 431) were uploaded, we were
ready to begin testing.

Using the University Testing Services online testing format has many
advantages for those wishing to assess university curricula. Online access allows students to complete the surveys when they are most motivated and at their convenience. However, researchers can remotely control when the surveys are available, and blocks of dates that the tests were open/accessible were designed to match the experimental design (see section 3.1.2). The website used for testing is run through the Penn State system, so students sign in using their Penn State Access Account ID (i.e. xyz123) and password. These are well known to the students since they are required for a wide range of university activities (including email). Thus by using this system, we avoided the logistical difficulties of assigning unique IDs and passwords for a restricted survey (lost IDs, forgotten passwords, etc.). When students sign in they are taken to a page that lists all courses that have surveys in this system. This assures that students who are not in the course being tested cannot take the NMR survey by accident. Responses can be monitored as they are received, and each completed survey was tagged with a time and date stamp. When all responses were received or when testing was closed, all data were downloaded into an Excel spreadsheet, and scoring and data analysis could begin.

Another benefit of using the UTS website is the control offered to researchers and/or faculty through the test maintenance page at the same website. Remote access is provided 24 hours per day, 7 days a week through this site. This section of the website is secure, and only the researcher (or instructor) can enter this section and make changes. The class lists for each course must be entered by the UTS staff, but additions to this list can be made individually by those conducting assessments. The procedures to modify or add surveys/assessments were fairly straightforward, so
that we could immediately fix any problems in the questions or layout as they were detected. There were many choices of item formats, including multiple choice, essay, and “select all that apply” items. The online testing format itself was slightly restricted in that the output from the website had to be either text or pictures (jpeg), and the input into the website by students had to be text or selection of choices. However, researchers were compensated for these hinderances by the convenience of set-up and data collection.

3.1.2 Experimental Design

For this assessment, we wanted to compare the new NMR curriculum to an alternative, and so a comparison group was needed. The experimental (E) group contained students who had completed the hands-on NMR module at the time of the posttest. The control (C) group contained students who had not completed the modules at the time of testing. Random assignment is an important feature of “true” experimental designs in educational research. Students are randomly assigned to an experimental condition (E or C group). Random assignment means that every participant has an equal chance of being placed into the E or C group, so that any differences in the two groups are a result of chance alone. This is done to control for the participant characteristics validity threat, in which the subjects of one group have some factor in common that makes them different from the other group. For example, imagine a design where an 8 a.m. lab section is chosen as the control group and a 2 p.m. lab section is chosen as the experimental group. Perhaps, the majority of students in the 8 a.m. lab registered late, and so have some common
factor governing their lab selection. If this was the case, then this group may have different participant characteristics than those who registered early and were able to obtain into the 2 p.m. lab section. To avoid this threat, random assignment was used for this study.

Random assignment was used within both courses being studied in this assessment. Chem 431 students only have one lab section, and so all students have the same instructor, teaching assistants (TAs), and lab meeting times. Chem 457 has multiple lab sections, and so while students in each section have the same instructor and access to help for the NMR modules, the sections vary in time of day, day of the week, and TAs. Rather than assigning groups by lab section, every student was randomly assigned to an experimental condition (E or C) so that every lab section tested had students from each group. This method of selection made the E and C groups essentially equivalent. It should be noted that while participation in the assessment (surveys) was optional, participation in the NMR lectures and hands-on experiments of the curriculum was not optional. All students in each course were randomly assigned to an experimental condition (E or C group) before testing began, but a different number from each group elected to participate in the assessment.

The general experimental design chosen for this assessment is shown in Figure 3.2. While testing schedules had to be modified to fit within confines of the individual course and NMR modules, the basic features for used for both courses are represented. As was previously described, a pretest and posttest were used, but not all variables were included on the pretest assessment (see section 3.1.1). The pretest
was given after the NMR lecture because we wanted to determine if the lecture in conjunction with the hands-on NMR module were superior to the lecture alone. This way we were testing both the hands-on and depth (through multiple exposures) features of this new NMR curriculum. Between the two testing periods, the E group completed the hands-on NMR experiment, and the C group performed an alternate experiment(s) that was unrelated and did not have NMR content. Though students in the control group had not completed the hands-on NMR experiment at the time of the posttest, they did receive the module after testing was completed. Since we believed that the NMR modules were beneficial to students, it would be unethical to deny some students access to this learning. By using this design, we could allow all students to participate in the experiments without impeding the assessment.

**Figure 3.2. Planned Testing Schedule for NMR Assessment.** The experimental design for this assessment required an experimental group (E) and a control group (C). Both experimental conditions received a brief re-introduction to NMR (NMR lecture), before the pretest was administered. The experimental group then completed the hands-on NMR module. The control group performed an experiment unrelated to NMR during this time period. Both groups were administered the posttest at the same time. Following this test, the control group could complete the NMR experiment without hindering the testing.
The experimental design for Chem 457 was very close to the planned design shown above, however slight modifications had to be made. The testing schedule used for this course is shown below (Figure 3.3). Testing was completed outside of class time, and so the online test was only open between the last class period of one activity and the beginning of another. The hands-on NMR module for this course was the $T_1$ experiment, and required students to spend approximately 45 minutes to set-up and run the experiment on the 400 MHz spectrometer. Students then spent 30 minutes processing the NMR data. To avoid overcrowding and inactive members of a group, students worked in pairs for this experiment. The number of students that could complete the $T_1$ experiment in a given laboratory period was limited to only four, two pairs per period. Pairs took turns working on the spectrometer for either the first or second half of the period. As shown below, group 1-4 were the experimental group, and they completed the hands-on NMR module while the control group worked on non-NMR related lab experiments. Then the posttest was administered. By having all participants take the posttest at the same point in the course, we could control for maturation, a validity threat that comes about when students’ scores improve because of natural changes that occur over time due to aging and experience. However, some subgroups of the experimental group completed the NMR module earlier than others meaning that the time between completion of the experiment to when the posttest was taken varied for the E group. If anything, this would disfavor the experimental group, and so any results found to be significant can still be trusted.
Figure 3.3. Testing Schedule Used for Chem 457. Due to constraints of the hands-on NMR module for this course, the following testing schedule was used for Chem 457. This represents the schedule for one lab section, but all lab sections followed the same schedule. This lab course meets once a week. All Chem 457 students received a lecture on NMR as their first exposure to the topic within this course. After the pretest, groups 1-4 completed the hands-on NMR module, which for this course was the T1 experiment. Each group consisted of two pairs of students who completed the experiment in either the first or second half of that lab period (four students total). After groups 1-4 completed the module, a posttest was given to all participants. Then, groups 5-7 completed the hands-on NMR module. In this setup, the experimental (E) group was made up of lab groups 1-4, and the control (C) group was made up of lab groups 5-7.

Chem 457

Hands-on NMR Module: T1 Experiment

<table>
<thead>
<tr>
<th>Completed Hands-on NMR Module</th>
<th>Class 2</th>
<th>Between Class 2 &amp; 3</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 6</th>
<th>Between Class 6 &amp; 7</th>
<th>Class 7</th>
<th>Class 8</th>
<th>Class 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>E or C group</td>
<td>- -</td>
<td>- -</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
<td>Group 4</td>
<td>- -</td>
<td>Group 5</td>
<td>Group 6</td>
<td>Group 7</td>
</tr>
<tr>
<td>Other Related Activity</td>
<td>NMR Lecture ALL</td>
<td>Pretest ALL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Posttest ALL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E = Experimental group  
C = Control group

The experimental design for Chem 431 required more compromises due to constraints of the course itself. This course’s reliance on NMR throughout the semester and many overlapping assignments meant that the only time that testing could take place without sizable interference was at the beginning of the semester. Rather than testing all students at once, the two groups had to be given the posttest at different times, since there were very few laboratory activities that the control
group could complete prior to learning the NMR software. The testing schedule for Chem 431 (Figure 3.4) contained many of the same features as those previously described. The re-introduction of NMR concepts for this course was accomplished through a written summary of relevant topics. Again, both assessments were completed outside of class time through the online testing system. After the pretest, several laboratory activities were completed that were unrelated to NMR. After the control group was given the posttest, both groups completed the new NMR activities: the XWIN tutorial and Analysis of a Crude Product assignment. The posttest was administered to the experimental group after the crude product assignment was due, but before work on the next assignment was begun.
Figure 3.4. Testing Schedule Used for Chem 431. Due to constraints of the course, the following testing schedule was used for Chem 431. Only one section of this course is offered per year, and lab meets twice per week. For this course, a re-introduction to NMR topics was given in written form (Intro NMR), and this was the first exposure to the topic within this course. After two class periods where students worked on lab activities not related to NMR (Non-NMR below), the control (C) group completed the posttest. Then all students completed the hands-on NMR module, which for this course was the XWIN software tutorial teaching them how to “work-up” their NMR spectra. Students then had several class periods to complete their “Analysis of a Crude Product” assignment (Crude Product). Upon completion of this assignment, the posttest was administered to the experimental (E) group.

Chem 431
Hands-on NMR: XWIN Tutorial & Analysis of Crude Product

<table>
<thead>
<tr>
<th>Class 2</th>
<th>Between Class 2 &amp; 3</th>
<th>Class 3</th>
<th>Between Class 4 &amp; 5</th>
<th>Class 5</th>
<th>Class 6</th>
<th>Class 7</th>
<th>Between Class 7 &amp; 8</th>
<th>Class 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab Activities</td>
<td>Intro NMR</td>
<td>-</td>
<td>Non-NMR</td>
<td>Non-NMR</td>
<td>-</td>
<td>XWIN tutorial</td>
<td>Crude Product</td>
<td>Crude Product</td>
</tr>
<tr>
<td>Testing E group</td>
<td>Pretest</td>
<td>-</td>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing C group</td>
<td>Pretest</td>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E group = Experimental group
C group = Control group

There were several validity threats using this testing schedule for Chem 431. The pretest and posttest are administered close together for the control group, and this represents a testing threat (test-retest). The concern is that participants’ scores will improve simply from taking the test more than once, and this is more likely for tests that are taken close together. However, administering a pretest had benefits (e.g. assuring that the E and C groups were equivalent) that outweighed a possible testing threat. A test-retest threat would favor the control group for variables covered on the pretest (especially depth of understanding and basic NMR
knowledge). However, problem solving and confidence for class skills would remain *unaffected* since these measures were excluded from the pretest. The experimental design for Chem 431 also contains a maturation threat, but this would favor the experimental group.

### 3.1.3 IRB Approval

Any research performed using human subjects at PSU, even surveys, must be approved by the Office of Research Protections (ORP) through an Institutional Review Board (IRB). Professors have been monitoring their own teaching without this formality for decades, but when it becomes “research,” different procedures must be followed. For the ORP at Penn State, the main distinction between instructional methods and educational research hinges upon dissemination. According to their site,\(^\text{64}\) “**Dissemination** includes, but is not limited to, honor's, master's, and doctoral theses; presentation at a scientific meeting or conference; submission to or publication, paper or electronic, in a scientific journal; and Internet postings.” Thus, any chemical educator planning to disseminate his/her results should become familiar with the processes of the IRB.

The goal of the IRB is to protect the rights of participants used in research and to make sure they are informed of any dangers associated with a research study. Penn State has a created a separate IRB for social sciences to protect participants in this type of research from undue stress or psychological harm. For example, we requested access to participants’ grades on certain assignments as part of the pilot study, and one review member warned that this could be very stressful for some
students (though we still received approval). Two more primary concerns of the IRB are *coercion* and *consent*. It is possible that if students think that lack of participation (or poor performance) on a survey may adversely affect their grade in a course, they will feel coerced. Thus, the instructor in the course should be distanced from the assessment process as much as possible. Teaching faculty should *not* be able to identify individual student’s responses or scores on testing instruments, and the list of those who participated in the assessment should be kept from them until the end of the semester so as not to introduce possible bias. Every participant in a study must give their *informed consent* to be included in the research. In this process, the researcher discloses all procedures and potential risks through a form, and the participant acknowledges that he/she understands these and that he/she is a willing participant by giving his/her signature. The approved consent form for this study can be found in Appendix D (Section 1).

All research materials and procedures were approved through the IRB at Penn State University (IRB # 21504). Not only must all testing instruments be submitted, but also any recruitment scripts and materials. Appendix D contains samples of the script used to introduce students in study 2 to the NMR assessment (section 2) and the handout they received (section 3). Some measures were modified after the pilot study and the procedures were slightly different for studies 1 and 2. For all changes, paperwork was submitted through the IRB to receive approval *prior* to use.

The most important procedural consideration for this study was the use of incentives. Since participation in the online surveys had to be voluntary, we felt
students should receive extra credit for their time spent outside of class. This is allowable provided two conditions are met. The first is that the total amount of extra credit offered cannot exceed 3% of the total class grade. Also, an alternative method of receiving extra credit that is equivalent (in time and effort) must be offered for students that still do not wish to participate in the study. For this assessment, students were offered the alternative of completing an article review for the same amount of extra credit, though no students utilized this option. After collecting data for study 1, researchers sought to increase participation in the study by offering another incentive. There are strict guidelines on payment of participants, however a prize drawing can have an equal effect for considerably less effort and cost. For study 2, we offered entry into a random drawing for an iPod Nano66 for students that completed both surveys. This was approved through the IRB. Indeed, the percentage of participants was higher for the second study (79 students out of 85 enrolled), and at the end of the semester one of them received the iPod through a random drawing containing the names of all who participated.

3.2 Pilot (Fall 2005 - Spring 2006)

After designing instruments and procedures for the NMR assessment, it was necessary to run a pilot study52 to determine if changes in testing measures or methods needed to be made. So far, this thesis has described the final testing procedures and materials used for data collection in studies 1 and 2 (Fall 2006 - Spring 2007). Therefore, we must first describe how methods used for the pilot study compared to the final studies.
The pilot study for this NMR curriculum assessment was conducted during the 2005-2006 Penn State academic year, and included responses from 108 students taking Chem 457 during this period. During this time, proper implementation of the testing procedures was the goal. We introduced our methods to teaching faculty, and worked on streamlining interactions with the online testing system. In general, teaching faculty participating in this program were very supportive of this work, but it took some time to smooth out the lecture and homework schedules. Also, there was initially some confusion regarding their level of involvement in assessment and how to avoid coercion of the students. Working with the online testing system also required some trouble-shooting so that all students who needed access were able to complete surveys.

The pilot study was used to weed out measures that were not suitable for the final assessment. One measure from the pilot study that concerned researchers was confidence for analytical instruments. For these items, students were asked “For each listed analytical technique listed, rate your level of confidence using the scale provided (1 - 5).” The analytical techniques provided were UV-Vis, fluorescence, GC, GCMS, IR, and NMR, and these items were far more general than the task confidence items previously described. The goal of the confidence for analytical instruments measure was to provide a context for NMR confidence among other techniques. Also, we believed that increased confidence in NMR might affect confidence for other techniques. Though there were moderate differences in pre-post levels of confidence for several of these techniques, it became clear that these could not be reasonably attributed to the NMR module, since students in both
experimental conditions (E and C groups) had worked with other analytical techniques during the same time period. There was no way to isolate their experiences with the many analytical instruments used in the course, and therefore this measure was eliminated prior to the final assessment.

A measure probing student attitude toward course-specific factors (e.g. lecturers, teaching assistants, etc.) was present on the initial assessment used in the pilot study, because these could vary and influence responses on other measures. Items for this measure were taken directly from the Chemistry Attitudes and Experiences Questionnaire (CAEQ). For these items, students were asked to use a five-point Likert scale to rate their attitude (interest, perceived access to help, etc.) on items referring to lectures, TAs, and content within the course. For this pilot study, students indicated that it was easy to access help from lecturers and TAs, but that material was sometimes uninteresting and concepts were not always presented clearly. These were general items meant to cover all course activities that took place between the testing periods, and it is logical to think that student’s feelings towards the course as a whole might affect their motivation, interest, and effort for NMR activities within this context. However, how does this relate to the research goals? Looking for a correlation between attitude toward the course and NMR outcomes seemed beyond the scope of this work, and so this measure was omitted from future studies.

The most important difference in the pilot study was the problem solving task, which was markedly different from the final HETCOR problem used. The initial measure of problem solving was an optional homework type assignment that
contained a number of physical chemistry calculation problems directly related to the lecture materials. While this measure had the benefit of having multiple items, there were a number of logistical concerns associated with this type of assignment. Administered as a written assignment, students took it home and worked on it over the course of a week. There was no need to work in groups for this problem set, since it was not graded (any student that completed it received the full extra credit points). However, for all other assignments in the course, students use outside texts and work in groups. On such a similar assignment, there was no way to ensure that students were not using the same methods to complete the work. While this is still a concern for the HETCOR problem, working together is logistically more difficult since it is administered on the computer. Also, the style of the HETCOR problem ("describe everything you know") was not as conducive to group work or use of reference materials. Another benefit of the HETCOR problem over the pilot problem solving measure was its link to another course being tested (Chem 431). The problems given on the pilot study were very content specific and could only be solved by students that had taken a physical chemistry course.

Pilot studies should also be used to clarify and improve upon even those procedures and measures that seem to be working. The same experimental design was used for Chem 457 in both the pilot study and studies 1 and 2, and many of the same measures were used as well. The latter included: depth of understanding, basic NMR knowledge, confidence for NMR and general science tasks. For these, only minor wording changes and removal of “bad” items were made before they were integrated into the final assessments. Several items were removed from the
confidence for general science tasks measure. For an example, whether or not students felt confident “determining the appropriate units for a result using a formula” seemed less relevant to learning NMR or an analytical technique than the other items in the measure, and so it was removed.

Through data collected on students’ prior exposure to NMR topics in the pilot study, researchers decided that this section could be streamlined. On the first version of the pretest, we asked students not only what courses they had taken (that contained NMR), but also what NMR concepts they had covered. During piloting, we found that students could not consistently recall even the general types of NMR information that had been covered in their courses. In many cases, students who had taken the same course did not list the same information. Therefore, the item asking for specific topics covered was dropped from the measure. The primary determinant of prior exposure in studies 1 and 2 became the courses taken, and researchers relied on courses’ syllabi and discussions with teaching faculty (and TAs) to confirm the topics covered.

3.3 Data Collection Methods for Studies 1 and 2 (Fall 2006 - Spring 2007)

3.3.1 Participants

Participation in this NMR assessment was open to all students enrolled in the courses being studied (Chem 457 and Chem 431) during the 2006 - 2007 school year at Penn State. Summaries of enrollment and participation data are shown
below (Figure 3.5). Both courses were offered in the fall, but there were only a moderate number of openings available. Chem 457 is also offered in the spring, and more students take this course during the spring semester. For study 1, participation was lower than expected even though the students were offered extra credit. One reason may have been students forgetting when the surveys were open. As evidence of this, every student participated in the pretest for Chem 431, when the survey was open the same day that the NMR assessment was introduced within the course. For study 2, email reminders were sent to the entire classlist to be sure that students knew when the surveys were open and closed. Also, the iPod drawing was added as an extra incentive, and these changes in recruitment methods may have been directly responsible for a higher percent of enrolled students participating. In the spring semester (study 2), the 79 students that did at least one survey represented 93% of the class enrollment (85 students), and 80% completed both surveys.
Figure 3.5. Participants for Studies 1 and 2. All students enrolled in Chem 457 and Chem 431 during the 2006-2007 school year were invited to participate in the NMR assessment studies. The breakdown of enrollment and participation is shown below.

<table>
<thead>
<tr>
<th>Study 1: Case 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chem 457 Fall 2006</td>
<td>Chem 457 Spring 2007</td>
</tr>
<tr>
<td><strong>n students</strong></td>
<td><strong>n students</strong></td>
</tr>
<tr>
<td>Openings</td>
<td>70</td>
</tr>
<tr>
<td>Enrolled</td>
<td>38*</td>
</tr>
<tr>
<td>Openings</td>
<td>120</td>
</tr>
<tr>
<td>Enrolled</td>
<td>85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>n students</strong></th>
<th><strong>% enrolled</strong>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>19</td>
</tr>
<tr>
<td>Posttest</td>
<td>19</td>
</tr>
<tr>
<td>1 Survey</td>
<td>21</td>
</tr>
<tr>
<td>Both Surveys</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>n students</strong></th>
<th><strong>% enrolled</strong>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>77</td>
</tr>
<tr>
<td>Posttest</td>
<td>70</td>
</tr>
<tr>
<td>1 Survey</td>
<td>79</td>
</tr>
<tr>
<td>Both Surveys</td>
<td>68</td>
</tr>
</tbody>
</table>

Study 1: Case 2
Chem 431 Fall 2006

<table>
<thead>
<tr>
<th><strong>n students</strong></th>
<th><strong>% enrolled</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Openings</td>
<td>75</td>
</tr>
<tr>
<td>Enrolled</td>
<td>46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>n students</strong></th>
<th><strong>% enrolled</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>46</td>
</tr>
<tr>
<td>Posttest</td>
<td>24</td>
</tr>
<tr>
<td>1 Survey</td>
<td>46</td>
</tr>
<tr>
<td>Both Surveys</td>
<td>24</td>
</tr>
</tbody>
</table>

* Estimated total enrollment based on researchers’ information. This number represents the minimum possible n, and maximum possible percent(%).

3.3.2 Instruments

The testing instruments used for studies 1 and 2 were administered as online surveys outside of class time (see section 3.1.1). These included measures of depth of understanding, basic NMR knowledge, problem solving (HETCOR problem), confidence for NMR tasks not performed as part of the course, confidence for class
skills used in the hands on NMR modules, and confidence for general science skills. Items types included free-response, multiple choice, and 1 to 7 point rating scales depending on the variable being measured. The development of these measures was described in Chapter 2 (section 2.3). The final versions of the assessments used for both Chem 457 and Chem 431 are given in Appendices B (posttest) and C (pretest).

3.3.3 Procedures

The testing schedules used for Chem 457 and Chem 431 are described in section 3.1.2. All testing procedures and materials were approved for use by Penn State’s IRB (#21504). See section 3.1.3 for further details.

3.4 Rubrics and Scoring

Data were collected using the methods described above, but before these could be analyzed, data had to be organized and scored. The online testing website used for this assessment allowed the researcher conducting the assessment to download all data for a given course into Excel. Though a copy of the participant’s identifying information and their experimental condition must be kept on file for later use (and kept confidential), this information should not be accessible during scoring. A new spreadsheet was created where all identifying information was removed and names were replaced with unique IDs. These IDs were generated using the random number generator in Excel, and care was taken to avoid using the same numbers for different data sets. ID numbers were assigned by participant not response, meaning that the pretest and posttest responses for the same person had
the same ID number. This is important because scores must be matched for analysis using the dependent $t$-test between pretest and posttest scores.

Responses were scored by the researcher, and in some cases, a trained scorer. All scoring was completed blind to condition, meaning that scorers did not know which responses belonged to the experimental group or the control group. Rubrics were developed for consistent scoring, especially of the free-response measures. For these measures, interrater reliability was tested and the results of this analysis is described in section 3.5.1. Only confidence measures did not require scoring since responses were given on a 7-point scale. After scoring was completed, item-analysis was calculated for all measures to gather information on each measure’s reliability (section 3.5.2).

A rubric is a tool used by scorers to consistently score items for a measure. Rubrics are also commonly used by practitioners in the field of education for scoring assignments and exams. The rubric for the basic NMR knowledge measure was straight-forward. On this measure answers were scored similar to an exam with answers that are either correct or incorrect. However, the depth of understanding and problem solving measures used only free-responses items, and so the range of responses was much broader. These would be considerably more difficult to score consistently without a good rubric.

### 3.4.1 Basic NMR Knowledge Rubric

The rubric used to score the basic NMR knowledge measure can be obtained by contacting the researchers responsible for this study. This measure consisted of
multiple choice and short answer items. For the reading spectra items, multiple-choice items were worth one point and short answer items were weighted (worth two points). All definition items were worth two points because they were deemed by NMR experts to be more cognitively involved than the reading spectra items. In this section, partial credit was awarded for answers that contained some but not all of the pertinent information, or that contained some incorrect information in addition to the correct answer.

3.4.2 Depth of Understanding Rubric

Depth of understanding was measured using the scores from two items, and each of these was scored using its own rating scale (for complete rubric, see Appendix E). These were designed to cover the two key facets of depth of understanding of NMR that were believed to be lacking from the majority of curricula found in the chemistry community. Depth of understanding, according to the authors of the project grant, required building upon and linking the main aspects of NMR (applications, theory, and instrumentation). The first item, depth Q1, was designed for students to convey their view of NMR as a whole (and those aspects deemed most relevant). Also, students with depth of understanding should see beyond the most obvious uses of the expensive analytical instrumentation in their labs. The complexity of NMR is rivaled only by the variety of chemical information that it can be used to obtain, and a deeper understanding must include an expanded view of its capabilities. The second item in depth of understanding, depth Q2, was designed to collect this type of information. By using a free response
format for these items, students were able to show what they considered most important, and their answers were not biased by the knowledge of the test makers. However, this format required the development of rubrics to consistently score items. The general process utilized when creating a rubric of this type can be found in Figure 3.6.

**Figure 3.6. Steps in Making a Rubric.**

<table>
<thead>
<tr>
<th>Making a Rubric for Open-ended Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Sort responses into discernable groups with a sequential hierarchy.</strong> This should be done using at least 20% of the responses given. The number of groups will vary depending on the data. Having more groups improves variation of answers and thus the likelihood of distinguishing between experimental conditions. However, this must be balanced with reliability. For example, if a scorer cannot distinguish between a 4 and 5 most of the time, then that distinction may be a false one.</td>
</tr>
<tr>
<td><strong>2. Create a detailed rubric.</strong> This will improve reliability in scoring for you, and other scorers as well. Also, this is a good time to think about whether training scorers will be necessary.</td>
</tr>
<tr>
<td><strong>3. Consult with an outside expert in the content area, if possible.</strong> A scale based on only one person’s opinion may elicit questions of validity. If you are able to discuss interpretations of responses and especially rankings with an expert in the field and come to an agreement, this will add to the validity of the rubric.</td>
</tr>
<tr>
<td><strong>4. Use rubric to score all responses.</strong> Scoring must be done blind to experimental condition.</td>
</tr>
<tr>
<td><strong>5. Have an outside scorer score responses to determine interrater reliability.</strong> Train outside scorer, if necessary. It is important to document all aspects of training. The outside scorer should score at least 20% of the responses given. Use correlation statistics* to determine the interrater reliability.</td>
</tr>
</tbody>
</table>

*Pearson’s $r \geq 80\%$ is used as a cut-off for acceptable interrater reliability.
When creating a rubric, it is preferable to have literature precedence to support how scoring groups are chosen. Work by Abraham\textsuperscript{68} provided a model for assessing students’ understanding of chemistry concepts that was found to be relevant to scoring of depth of understanding items. In this study, students were prompted with specific chemistry situations designed to illustrate a certain chemistry concept, and given a mix of closed-ended and open-ended questions relating to this prompt. The levels of understanding displayed by students’ responses were rated on a 6 point scale (Figure 3.7). The scale ranges from No Understanding (NU) to Sound Understanding (SU), and the rank of Partial Understanding (PU) was reserved for answers between these extremes. A rubric was used to clearly define levels of understanding for each item in Abraham’s study. Also, it is important to note that the presence of Misconceptions (M) was also a factor in Abraham’s evaluation of understanding. These affected students’ ratings as shown.

**Figure 3.7. Abraham’s Scale of Understanding.** The following scale was developed and used by Abraham\textsuperscript{68} to evaluate students’ understanding of chemistry concepts. A rubric outlined the level of understanding represented by specific responses for each item in Abraham’s study. Misconceptions were also noted, and they contributed to students’ final ratings for a given item.

<table>
<thead>
<tr>
<th>NR</th>
<th>NU</th>
<th>M</th>
<th>PU/M</th>
<th>PU</th>
<th>SU</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Response</td>
<td>No Understanding</td>
<td>Misconceptions</td>
<td>Partial Understanding/Misconceptions</td>
<td>Partial Understanding</td>
<td>Sound Understanding</td>
</tr>
</tbody>
</table>
Requirements for a response scored as partial versus sound understanding were clearly defined in the rubric for each chemistry concept measure. Using this scale, Abraham was trying to determine how a student’s age level (junior high, high school, or college aged) and/or reasoning ability might affect his/her understanding and misconceptions of specific chemistry concepts. Several general trends were found using this scoring method.

For the first item (depth Q1, “In general, what is NMR? What does it do?”), there are several correct types of answers. A perfect answer to depth Q1 would cover each of the following categories of information: magnetic, purpose, spins, and process. NMR is defined as nuclear magnetic resonance and it is a technique that utilizes the magnetic properties of nuclei. These ideas are included in the “magnetic” category. This technique relies on population differences of nuclei in opposite spin states, and during an NMR experiment these spins are manipulated. The “spin” category is used to describe answers that portray this aspect of NMR. Some students spoke of the uses of NMR to find structure or other information about a chemical species, and these types of responses were included under the “purpose” heading. The “process” category is needed for responses that contain steps of an NMR experiment and what is happening during this process. Answers that discuss the types of pulses, vector descriptions of the net magnetization during pulses and delays, and processing such as Fourier transform would all be included under the “process” heading.

When scoring depth Q1, correct statements matching any of the four categories contributed to the score for student’s level of understanding. An example
of the scoring sheet used to score this item is shown in Figure 3.8. Each of the four categories mentioned answered some aspect of depth Q1, but no category was considered better or more important than any other category. Indeed, for students to have a sound understanding of NMR, they needed to understand multiple aspects of this technique. For this reason, the number of categories that contain correct statements (# C category) determined the student’s rating for this item. Correct statements within the same category were not counted twice, so that a person with three correct statements in the “process” category was rated lower than a student with one correct statement in both the magnetic and spins categories. For this item, it was also important to note misconceptions held by the students. A student whose answer contained misconceptions was be rated lower than one whose answer did not. For simplicity, the final rating did not consider the number (or severity) of misconceptions, only their presence or lack thereof.

Figure 3.8. Sample from scorer’s sheet for depth Q1. The first three columns are used for matching the scoring to the sheet of written responses. The next four columns contain the four key categories for responses to depth Q1. A check in these boxes corresponds to at least one correct response in that category. The sum of these checks is recorded under “# C Category” (possible scores, 0–4). The presence of misconceptions is recorded under “# M”. The latter two columns (# C and #M) are used to determine the final rating for this item.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>ID</th>
<th>Depth Q1</th>
<th>magnetic</th>
<th>purpose</th>
<th>spins</th>
<th>process</th>
<th># C category</th>
<th>#M</th>
<th>depth1 rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>57S07</td>
<td>0.08182</td>
<td>Nuclear magn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57S07</td>
<td>0.20734</td>
<td>It is a method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The rating scale for depth Q1 is shown below (Figure 3.9). The lowest score (0) contained low-level responders, those who gave no response, those with
misconceptions (all M) and those with vague answers that did not refer to any of the previously described categories. An answer with three or more categories (and no misconceptions) was considered at the sound understanding level. Students who had stated misconceptions were ranked lower than those without, however there was one exception. A few responses seemed to be at the sound understanding level, but still had stated misconceptions. It is possible that higher level students showed a tendency to write more, leading to more insight into their thought processes. Indeed, consultation with an expert in NMR confirmed that the “3+ correct + M” responses exhibited a greater grasp of the technique than was shown by responses that covered only one of these categories (1 correct).

**Figure 3.9. Rating Scale for depth Q1.** In this six-point scale, the number correct refers to the number of correct categories that are discussed in a student’s response. An M represents one or more misconceptions that were present in a particular response.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>vague answer (no correct)</td>
<td>1 correct + M</td>
<td>2 correct + M</td>
<td><strong>1 correct</strong></td>
<td>2 correct</td>
<td><strong>3+ correct + M</strong></td>
<td>n/r</td>
</tr>
<tr>
<td>all M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n/r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 3+ correct = 3 or MORE correct

The second item for this construct, depth Q2, focuses on a different aspect of NMR understanding. Here, we were interested in what students knew about NMR’s capabilities to elucidate chemical phenomena. The question “what types of information about a molecule/reaction can you get using the NMR spectrometer?” could have many possible correct answers, and many examples are listed in the scoring instructions (Appendix E). As with depth Q1, there is no real hierarchy to
the answers given. For example, knowing that NMR can be used to find the conformation of a molecule is no more important than knowing that NMR can be used to measure diffusion in a solution. However, a simple count of all correct answers was not sufficient either because students sometimes responded with uses that were found using the same type of analysis (e.g., percent yield and reaction completion). To account for this, equivalent answers such as this were denoted by ‘=’ sign on the rubric, and each type was counted only once.

The scoring of depth Q2 was unique in that misconceptions were not considered. The reasons for this are two-fold. One reason is that for this item it was more difficult to determine whether there is a misconception. In some cases, it was unclear whether students thought that the chemical information was found through a direct measurement. For example, while it may be possible to determine the molecular weight of a compound indirectly using NMR (once the structure is obtained), this is not the main purpose of the experiment, whereas a mass spectrometer directly measures molecular weight. In other cases, it was difficult to know if there was a misconception because NMR is so versatile that for some information (such as density), an expert could design a complicated method to gather this information using NMR. However, it seemed unlikely that an undergraduate population, which is expected to contain mostly (if not all) novices, would understand how this could be accomplished. For these reasons, responses such as these were not counted as correct, but neither were students penalized for answering in this way. Another problem that could arise from counting misconceptions was that depth Q2 was designed to elicit multiple answers (“as many
as possible”). Students who listed more answers were writing more, and thus were more likely to reveal any misconceptions they had. Students who listed only one type of information would not have this problem. By not counting misconceptions, we avoided giving the latter an unfair advantage in scoring.

Figure 3.10 shows the rating scale for depth Q2. The lowest score (0) represented a response with no correct information. This included those who gave no response, those with vague answers (e.g., reaction information), and irrelevant answers that did not answer the given question (e.g. shift). Students who received scores of 1 on this item were most often listing structure information. Since it was possible that even three answers could represent only experiments that students have seen in their courses (e.g. structure, purity, and T<sub>1</sub>), the ideal answer should be contain at least one more unique type of chemical information. And so, an answer with four or more correct responses was considered the maximum understanding level for depth Q2.

**Figure 3.10. Rating Scale for depth Q2.** In this five-point scale, a number correct refers to the number of correct types of chemical information that are listed in a student’s response. Samples of correct and equivalent responses can be found in Appendix E. Misconceptions were not counted for this item.

<table>
<thead>
<tr>
<th>0 correct</th>
<th>1 correct</th>
<th>2 correct</th>
<th>3 correct</th>
<th>4 or more correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### 3.4.3 Problem Solving Rubric

The problem solving task designed for this study consisted of a series of free-response items about a given HETCOR spectrum. To fully utilize all of the
information given in a HETCOR spectrum, students must use not only their knowledge of 1D NMR spectra but also synthesis of this information to eliminate wrong choices and arrive at the correct structure. We were interested in the process of problem solving, as well as the final answer. The open-ended question (hetcor 1), “using the HETCOR spectrum above, please list all of the information you can figure out about the molecule,” invited students to give write more than just a final guess. In addition, several questions were provided as scaffolding for the students, to help students make connections between what they have already learned and this novel task. These were not included in the final scoring, because any inspirations derived from this section would show up in hetcor 2, where students were asked to provide any “additional information” on the same HETCOR spectrum that they may have determined by relating the novel problem to their previous knowledge. When responses were scored for this problem solving task, both hetcor 1 and hetcor 2 were considered as two parts of the same response, and these will hereby be referred to as the HETCOR problem.

As expected, answers to the HETCOR question (parts one and two) varied from “no response” to listings of pieces of information (such as functional groups). Some students gave details about how they arrived at decisions (using splitting, shift, etc.), however not a single student actually came to a final conclusion about the identity of the molecule (ethyl crotonate, a.k.a. ethyl 2-butenoate). While there were obvious limitations in the online format used for testing (e.g. no drawing capabilities), not even one student gave the name of even an incorrect compound. It is believed that students were unable to answer the question, due to a difficulty level
above their normal capabilities (or possibly decreased motivation to engage in such a taxing task, considering the limited rewards/accountability). Although no students were able to solve the HETCOR problem, insights into their problem solving capacities were nonetheless available. In a study by Asieba and Egbugara,\(^1\) problem solving was measured using the individual steps required to solve a general problem. These steps were taken from Ashmore’s work with problem solving networks in chemistry.\(^6\) As shown in Figure 3.11, these steps are definition of the problem (i), selection of the appropriate information (ii), combination of separate pieces of information (iii), and evaluation of the solution (iv). Though these steps were developed in the context of solving chemical problems, they could easily be generalized to other problems in other content areas. By defining the necessary steps for solving a chemical problem, Ashmore aimed to aid undergraduates in the process by showing students how to relate pieces of information into networks (for step iii). The focus of work by Asieba and Ebugara differed in that networks were not used, but the general steps were used as a problem solving strategy for high school students. Not only was Ashmore’s model used as a guide to solving problems, but it was also used to assess the problem solving skills of the participants. The assessment was comprised of four quantitative chemistry problems (content was chosen to match that of the instruction given as part of the study), and was evaluated by giving a score for each step (i-iv) of the problem solving process.
Figure 3.11. Steps for Solving a General Problem. The following are the steps of problem solving as defined by Ashmore. Authors applied these steps to the training of undergraduates to use networks to solve problems in chemistry. Subsequently Asieba and Egbugara used this model to assess high school students problem solving skills.

<table>
<thead>
<tr>
<th></th>
<th>Definition of the Problem. Read the problem carefully. Determine the objective of the question. Subdivide into smaller problems to be answered first, if necessary.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ii</td>
<td>Selection of the appropriate information. Determine the relevant pieces of information needed to solve the problem. These may be given as part of the question, and/or through other resources (e.g., the periodic table or conversion charts).</td>
</tr>
<tr>
<td>iii</td>
<td>Combination of separate pieces of information. Combine smaller pieces of information through calculations or explanations.</td>
</tr>
<tr>
<td>iv</td>
<td>Evaluation of the solution. Check to see that the solution answers the stated question. Also check that the solution is reasonable (scale, units, etc.).</td>
</tr>
</tbody>
</table>

Though students in the current study were not explicitly taught to use Ashmore’s strategy for solving problems, use of at least some steps of process were evident in their responses. Students were asked to give any information that they could about the HETCOR spectrum given, and so gathering of relevant information (step ii) and combination of evidence (step iii) were necessary steps for well-formed responses. Definition and evaluation steps (i and iv) were not explicitly asked for, and while students may be using these steps in their thinking, it was unlikely that we would see them as written responses. Therefore, a rubric focusing on the intermediate steps (ii and iii) was constructed.

NMR is an analytical technique that can yield many types of information
about a molecule from a single spectrum, and the types of information given in the student responses can be arranged into a hierarchy. Figure 3.12 shows the coding key used for the problem solving rubric (Appendix F). The lowest level of information would be that taught to beginners in sophomore chemistry (lectures and labs). Here students learn what a peak represents (Th), how a peak’s shift relates to neighboring functional groups (Fg), how to obtain the determine the number of hydrogens per peak using integrals (#H), and what splitting means (Sp). These basic pieces of information represented one level using this scoring scheme. A student need not have every relevant piece of information to be counted in this category, but he/she must have at least one correct. Students learn the general ranges of shifts relating to functional groups in a molecule early on, however correct identification of certain functional groups within ethyl crotonate (the molecule used for the HETCOR task) would require more analysis and so were counted in a different category (see below).
Figure 3.12. Categories of Information Used in HETCOR Problem solving Task. This is an excerpt taken from the scorer's sheet for the problem solving task (Appendix F). Headings used for the rubric are explained and suggestions for note-keeping are given. Correct answers for some categories are also written here to serve as a reminder for the scorer.

<table>
<thead>
<tr>
<th># M number of misconception</th>
<th>Connected? or Combining</th>
<th>Info</th>
<th>Th number of types of hydrogen</th>
<th># H unique types of hydrogens</th>
<th>#c number of hydrogens (integral)</th>
<th>Types C &amp; number carbons, same answer</th>
<th>Sp splitting</th>
<th>Fg functional groups</th>
<th>RATING your rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>any #M (1 or more) will be counted the same</td>
<td>write &quot;connected&quot; or type of combining (&quot;ester&quot; or &quot;alkene&quot;)</td>
<td>mark as + or -</td>
<td>mark as + or -</td>
<td>mark as + or -</td>
<td>mark as + or -</td>
<td>mark as + or -; may be multiple answers, record all</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CORRECT ANSWERS > 5

**Th**
Number of types of hydrogen (H). How many unique hydrogens there are on a molecule. The number of peaks.

**# H**
Number of total hydrogens. The relative number of hydrogens per group is determined by using integrals.

**Fg**
Functional groups present in the molecule. By using the shift (ppm) of the peak, one can determine its proximity to an electronegative or shielding group. Many functional groups have characteristic shift ranges. (Shifting is often additive, so that a H that is alpha to a carbonyl and also neighboring an oxygen would be shifted further downfield that either group alone.)

**Sp**
Splitting is a result of $J$ coupling among hydrogen atoms on neighboring carbons (or other backbone atoms). By using the $n + 1$ rule, one can match hydrogens which are next to each other on a molecule.

**Tc = #C**
Number of types of carbon (C) and the total number of carbons are the same for this molecule. Integrals are rarely used in C-NMR because only symmetrical carbon chains lead to multiple carbons with the exact same shift.
**Connected**
A statement that indicates that students are using the 2D portion of the HETCOR spectrum, where each point represents a carbon connected to a hydrogen. This could be expressed as any of the following: shows 5 C’s with H’s attached, shows C and H that are attached, 5 sets of H and C are coupled, one C with no H’s attached.

**Combining Info**
A statement that indicated that students are combining pieces of information (iii of Ashmore’s steps). This may also be indicated by eliminating a possible choice (denoted by “NOT”). For example, if a student says “it is an ester or ether,” no elimination has been done and so they are not combining information. However, if a student were to write “ester” only, that would be evidence of combining multiple pieces of information (which would exclude ether as a possibility). Evidence of combining information could be expressed as either: ester (NOT ether), alkene (NOT aromatic)

The hierarchy of this information was discussed among NMR experts and a scoring system was assigned (Figure 3.13). Again, information that was at the sophomore organic level (Fg, H, and Sp) was at the lowest level of information. For this molecule, the number of unique types of carbon (C) and the number of total carbons (#C) are equal, and this represented an extension because early training focuses only on H-NMR. Even students who have never seen a carbon NMR could transfer some of the skills from H-NMR, but this type of transfer is usually difficult for beginners because the two spectra are quite dissimilar. Connecting information was considered at a higher level due to its particular importance for the HETCOR. This provided evidence of students using the 2D portion of the spectrum, and shows that they were including a type of information outside the realm of what they normally encounter (and have practiced). Combining information was considered equally important as it represents step iii of Ashmore’s problem solving process.
**Figure 3.13. HETCOR Problem Scoring.** The following shows the rating scale used for scoring the HETCOR problem and an excerpt from the scoring sheet used. This problem was scored on a 0-7 point scale, where higher numbers indicate higher levels of cognitive involvement on the problem. However, misconceptions (M) were also taken into account. The Scorer’s sheet was designed to help scorers keep track of scoring components. The ID and response columns were used to match the full response to the correct line of the scoring sheet. One or more checks in the M column mean that the response would be immediately put into the lower-level scoring bracket (1-3), and a response with no M would be scored using the higher-level bracket (4-7). For connected and combining info, the type must be written because having only one type results in a lower score than having both types.

Taken from Problem Solving Rubric

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Response</td>
<td>All M</td>
<td>Fg+H+M = #C +M</td>
<td>combining info (OR connected) + M</td>
<td>FG + #H</td>
<td>#C=Tc</td>
<td>combining info OR connected</td>
<td>combining info AND connection</td>
</tr>
</tbody>
</table>

M = misconceptions (1 or more)

7 = combining info. AND connected - write which ones (connected, ester, and/or alkene)
6 = combining info. OR connected - write which ones (connected, ester, or alkene)
5 = number of unique C’s - this is not information they learned in Chem 36, inferring that (number of peaks)
means the same as it would for H spectra
4 = correct functional groups, types/number of C, and/or types/number of H - chem. 36 level
   - lower levels all contain at least one misconception (M) -
3 = combining info. OR connected + M - still getting something beyond Chem 36 level
2 = functional groups, types/number of C, and/or types/number of H + M
1 = all misconceptions (all wrong answers) or nothing relevant, but they gave a response
0 = no response, didn’t try at all

**Taken from Scorer’s sheet**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>ID</th>
<th>response (1)</th>
<th>(part 2)</th>
<th># M</th>
<th>Combining Info</th>
<th>Connected?</th>
<th>answer</th>
<th>answer</th>
<th>Sp = splitting</th>
<th>Fg = functional groups</th>
<th>RATING</th>
</tr>
</thead>
<tbody>
<tr>
<td>57S07</td>
<td>0.08182</td>
<td>There are 6 ty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57S07</td>
<td>0.20734</td>
<td>There are five</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the cases described above, the corresponding levels of information are only counted if the information is *correct*, as misconceptions are important in this scoring. A student who reports an *incorrect* number of carbon’s would not have “#C”
contribute to their score at all. Also, if there are misconceptions in addition to correct information, this may count against the student. Using the ranking scale provided (Figure 3.13), one or more misconceptions (M) placed a response into the lower scores bracket (1-3), and having no misconceptions resulted in a score of 4 or higher. Some types of information collapsed into the same scoring category (denoted by a “+”), but this does not mean that all those types are required for this score. For example, a response that contained functional group (Fg) information but no information on the types of hydrogen (Th) was still given a score of 4. Likewise having more than one of these types of information (denoted by a “+” did not increase the score received. For example, a response that contained a misconception, the number of carbons (#C), and number of hydrogens (#H) would only receive a score of 2. The only exception to these statements was when a student had both connected and combining information in his or her answer. This response would receive a 7 (as denoted by the “AND” for this rank), assuming there were no misconceptions. If there was also a misconception, then a score of 3 would be given.

The scorer’s sheet for the HETCOR problem was designed to help keep track of scoring components (Figure 3.13). The first two columns (ID and response 1) were used to help the scorer match responses to the correct line of the scoring sheet. The rest of the columns represent decision points. If there is one (or more) check in the M column, then the response will be scored in the lower-level scoring bracket (1-3). The highest-level response (7) should have both connected and combining information components, thus it was important to denote the type (connected or combining info) on the scorer’s sheet. The scorer’s sheet readout also makes double-checking scores
easier, which may improve reliability.

3.5 Reliability Analysis

3.5.1 Interrater Reliability & Scorer Training

Rubrics are used to make scoring consistent, however open-ended responses are often subject to interpretation. Due to the subjectivity of any human scorer, interrater reliability must be assessed. In this process, scorer A scores a set of responses using a rubric. Then scorer B scores the same set of responses with the same rubric, and the scores are compared by calculating the Pearson’s r correlation between the two data sets. Depending on the nature of the responses and rubric, it may be deemed necessary to train scorers to use the rubric. For this assessment, items were first scored by the researcher (scorer A), and then another scorer outside of the research team (scorer B) was trained before scoring the set of scores. Scoring was always completed blind to condition.

Training Scorers

The first step of the training process was to choose a second scorer and schedule training sessions. For these items, it was important that the scorer have a sufficient background knowledge of NMR. However, we did not want to choose anyone who already had a connection to the project or our research group. Therefore, it was necessary to offer monetary compensation for the scorer’s time. The scorer recruited, Josh Sokoloski, was a Penn State graduate student whose research relates to NMR, but was unlinked to the Mueller group. Training occurred over two
sessions, one for the HETCOR problem (session 1) and one for depth of understanding items (session 2). All responses for the HETCOR problem were scored before the session 2 was begun. The HETCOR problem was arguably the more taxing to score, and this was why training and scoring of these responses were completed first. Each training session lasted for *no more than 30 minutes*.

Training was completed according to the process shown in Figure 3.14. Since the data in spreadsheet format, responses for all three data sets (study 1, cases 1 and 2, and study 2) were easily combined and printed out on a sheet (response sheet) that contained only the participant IDs and their corresponding responses. The scorer was also provided with a rubric and scorer’s sheet. The latter was used to record scores, and the order of matching ID’s was the same order found on the “response sheet”. The format used for scorer’s sheets can be found in the rubric section for the corresponding item type (HETCOR problem, section 3.4.3; depth Q1 and Q2, section 3.4.2). In session 1, the scorer was given a brief introduction to the purpose of the project and assessment. The purpose of each free response item to be scored in that session was described, and this was followed by a discussion of the rubric (including rating scale) and scorer’s sheet. Scorer B was allowed to ask questions at any point during this time, and indeed a two-way conversation about the scoring was achieved during both sessions.
Figure 3.14. Training Method for Scorers. The following method was used for training scorer B in this assessment. The same basic procedures were followed for all free-response items (HETCOR problem, depth Q1, and depth Q2). All scoring was done blind to condition (E or C).

<table>
<thead>
<tr>
<th>Training Procedure for Scorer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Training took place on two separate occasions. Session 1 consisted of training for the HETCOR problem. In session 2, the scorer was trained to score depth Q1 and depth Q2. Session 2 did not begin until after all scoring for the previous item (HETCOR problem) had been completed. Each training session lasted no more than 30 minutes.</td>
</tr>
<tr>
<td>2. For each session, Scorer B was provided with a printout of a responses sheet containing only the participant ID’s and corresponding responses (study 1 and study 2 responses combined). The rubric and scorer’s sheet were also provided.</td>
</tr>
<tr>
<td>3. A brief introduction to the purpose of the overall assessment was given. This was only required for session 1.</td>
</tr>
<tr>
<td>4. The free-response item(s) and its purpose was described. This was followed by discussion on the rubric and rating scale. Throughout this period, the scorer being trained could ask questions at any time.</td>
</tr>
<tr>
<td>5. Several items were scored together as examples. These were randomly selected prior to the training session. For each item, x responses were scored as examples by scorer A, and y responses were scored by scorer B and then checked against scorer A’s scores. (HETCOR: x = 5, y = 5; depth Q1: x = 5, y = 18; depth Q2: x = 5, y = 41.) Any differences in scores between scorer A and B were discussed.</td>
</tr>
<tr>
<td>6. For session 1 (HETCOR problem) only: Scorer B scored 20% (21 responses) on his own, outside of the training session. These were returned to scorer A, and compared against her scores. Differences in the scores were discussed.</td>
</tr>
<tr>
<td>7. All remaining responses for the item(s) were scored by scorer B on his own.</td>
</tr>
<tr>
<td>8. Pearson’s r correlation was calculated for each scorer’s scores (A and B) for each item using only the responses that scorer B had scored independently and that had not been checked with scorer A’s scores as part of training.</td>
</tr>
</tbody>
</table>
The bulk of the training session was used to practice scoring (Figure 3.14, steps 5-7). Responses were randomly selected to be used for this training, prior to the session. Scorer A (the researcher) gave examples of scores for several responses. Then, scorer B was given responses to score on his own. These were checked against scorer A’s scores within the session, and differences were discussed to make sure that a consensus was reached. For the HETCOR problem, 5 responses were scored as examples, and 5 more were checked. Then, scorer B scored an additional 20% of the responses (step 6) on his own outside of the training session. These were checked with scorer A’s scores for the same responses (IDs), and differences were discussed to ensure agreement. For the depth items (session 2), more responses were checked during the training session (step 5). This was possible because scoring of these items can be completed much more quickly. For depth of understanding items, the scoring and checking process was repeated several times until the number of responses ($y$) was reached (depth Q1: $y = 18$; depth Q2: $y = 41$). After training was completed, scorer B took all remaining responses and scored them independently (on his own time). Only these responses (not used for training) were used to calculate the interrater reliability (Figure 3.14, step 8).

**Interrater Reliability Calculations**

Pearson’s $r$ correlations were calculated to determine interrater reliability for free-response items using the Statistical Package for the Social Sciences (SPSS) program, and these results are shown below (Figure 3.15). Even though the number of total responses for each item was the same ($n = 113$), only responses not
used for training were used for Pearson’s $r$ calculations. This is why $N$ of responses used for the calculation varies for each item. For depth total, depth Q1 and Q2 scores from the same participants (ID) were added. As shown by the Pearson’s $r$ calculations, correlations between the scores of scorer A and scorer B were significant ($p = 0.000$) for all items. Correlations of $r > 0.8$ are considered acceptable for interreliability, and calculations for all items were found to be above this cut-off. Therefore, we can be confident that scoring methods using rubrics for these free-responses items (depth of understanding and problem solving) were reliable, and could be used by anyone with a sufficient background knowledge in NMR.
Figure 3.15. Interrater Reliability. Information used for calculating interrater reliability is shown below. The number and percent of responses scored independently (outside of training) by scorer B are reported. Only these scores were used for interrater reliability calculations, which were calculated as Pearson’s $r$ correlations between scorer A and scorer B scores. All items were found to be significantly correlated and showed acceptable levels of reliability ($r > 0.8$).

Responses Scored for Training vs. Independently.

<table>
<thead>
<tr>
<th></th>
<th>n scored for training</th>
<th>N scored independently</th>
<th>% scored for training</th>
<th>% scored independently</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Understanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth Q1</td>
<td>23</td>
<td>90</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Depth Q2</td>
<td>46</td>
<td>67</td>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>Depth Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Problem-Solving      |                        |                        |                       |                        |
| HETCOR Problem       | 35                    | 78                     | 22                    | 31                     |

* Added Q1 and Q2 for same ID to get total

Interrater Reliability Calculations

<table>
<thead>
<tr>
<th></th>
<th>Pearson’s r</th>
<th>Significance</th>
<th>N</th>
<th>Scorer A Mean (sd)</th>
<th>Scorer B Mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Understanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth Q1</td>
<td>0.820</td>
<td>$p = 0.000^{***}$</td>
<td>90</td>
<td>3.06 (1.23)</td>
<td>2.98 (1.32)</td>
</tr>
<tr>
<td>Depth Q2</td>
<td>0.877</td>
<td>$p = 0.000^{***}$</td>
<td>67</td>
<td>1.69 (0.874)</td>
<td>1.63 (0.813)</td>
</tr>
<tr>
<td>Depth Total</td>
<td>0.830</td>
<td>$p = 0.000^{***}$</td>
<td>67</td>
<td>4.72 (1.61)</td>
<td>4.54 (1.65)</td>
</tr>
</tbody>
</table>

| Problem-Solving      |             |              |       |                     |                     |
| HETCOR Problem       | 0.972       | $p = 0.000^{***}$ | 78    | 2.50 (1.79)         | 2.42 (1.77)         |

For Depth Total Q1 and Q2 scores that came from the same participant (ID) were added.

N = number of responses scored independently by scorer B

*** Correlation significant to the 0.001 level (2-tailed)
3.5.2 Item analysis

Once data were collected, we calculated the reliability of items within each measure. Cronbach’s alpha is an indicator of the internal consistency of items designed to measure the same construct.\textsuperscript{52} It is used by the researcher to determine if all of the items within a testing instrument are measuring the same variable/construct. However, it should be noted that this calculation \textit{cannot} confirm that the test is measuring the \textit{correct} construct, because that would be a matter of validity not reliability. For example, even if it is quite reliable, you cannot tell if the instrument made up of confidence for NMR tasks items are actually measuring the stated variable, task confidence. When using measures found in the literature, reliability already has been confirmed in most cases. However, this assessment required the development of new measures that we created, and so it is especially important that reliability be analyzed for these.

Item analysis was used to judge the reliability of each measure used for the NMR assessment (Figure 3.16). All calculations were performed using SPSS. Cronbach’s alpha can only be calculated for measures that contain \textit{more than} two items. For depth of understanding, there were only two items and so instead the Pearson’s correlation (Pearson’s \textit{r}) was calculated for this variable. Item-analysis cannot be performed on tests that contain only one item, and so the HETCOR problem was excluded from this analysis. Since we were trying to ascertain the consistency of items to measure the same thing, data sets from studies 1 and 2 were combined (study 1: Chem 457 and Chem 431 Fall 2006; study 2: Chem 457 Spring 2007) to assure that measures were reliable for all student samples (\( n = 113 \)). The
confidence for class skills measures had different sets of items for different courses, and so responses were grouped by course for these calculation (Chem 457: n = 89; Chem 431: n = 24).

**Figure 3.16. Item Analysis and Reliability.** Cronbach’s alpha was calculated for most measures. The values shown were calculated using all items for that measure. Both studies (study 1, cases 1 and 2, and study 2) were combined for reliability calculations. This means that n = 113 for most measures. Confidence for class skills was an exception since Chem 431 utilized a different set of items. Since there were only two items in the depth of understanding measure, a Pearson’s correlation was calculated instead of Cronbach’s alpha.

### Reliability of Items
*Data Sets Combined: Studies 1 and 2*

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Cronbach’s Alpha</th>
<th># Items</th>
<th># Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NMR Knowledge:</strong></td>
<td>Cronbach’s Alpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td><em>-0.034</em></td>
<td>2</td>
<td>113</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>0.629</td>
<td>13</td>
<td>113</td>
</tr>
<tr>
<td><strong>Problem-Solving:</strong></td>
<td>Cronbach’s Alpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>N/A</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Confidence:</strong></td>
<td>Cronbach’s Alpha</td>
<td># Items</td>
<td># Responses</td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>0.862</td>
<td>5</td>
<td>113</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td><strong>0.947 (Chem457)</strong></td>
<td>9</td>
<td>89</td>
</tr>
<tr>
<td>0.830 (Chem431)</td>
<td>10</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>0.872</td>
<td>6</td>
<td>113</td>
</tr>
</tbody>
</table>

N/A indicates that there were not enough items to justify reliability calculation

*Pearson’s correlation (r) of the two items gave this value, however it was non-significant (p = 0.722)

**Chem 457 courses only (both S07 & F06 data)**

The results of item analysis for measures used in the NMR curriculum assessment indicated high reliability for most variables. A Cronbach’s alpha of at
least 0.8 indicates good item reliability, and the calculated values of most measures were above this cut-off. When considering measures created by the researchers, all three task confidence measures were found to have good reliability (conf NMR tasks, \( \alpha = 0.862 \); conf class skills (Chem 457), \( \alpha = 0.947 \); conf class skills (Chem 431), \( \alpha = 0.830 \)). When the cut-off of Cronbach's alpha is not met, it is often possible to improve reliability by removing items that have low or negative correlations to the other items in the test. This procedure was followed for the basic NMR knowledge measure (all items, \( \alpha = 0.629 \)). It was found that by removing items 1 and 2 from the reading spectra section (rs1 and rs2), a maximum value of Cronbach's alpha (\( \alpha = 0.645 \)) could be reached. Looking more closely at the responses for rs1 and rs2, it became obvious why these two items were contributing negatively to Cronbach's alpha calculations. Both items showed very little variance among the participants, and most students got full credit (one point) for rs1 and rs2.

Item analysis for the depth of understanding measure was very revealing. For this measure, the Pearson's correlation among these two items was low and not significant (\( r = 0.034, p = 0.722 \)) indicating no correlation. Recall that for the depth of understanding measure, the researcher responsible for assessment and NMR experts were trying to measure two distinct aspects of depth of understanding, students' understanding of the different aspects of NMR and their knowledge of the variety of information that NMR can give. Correlation should not be expected if these aspects do not overlap. In this case, lack of correlation may be an indicator of discriminant validity. A measure that has **discriminant validity** gives results that are different from a set of items measuring a different variable or construct. By
adding more items that measure each aspect of depth and performing another item analysis, we could determine whether or not our depth of understanding items were measuring different constructs. Items measuring the same facet of depth of understanding should correlate, and those measuring different aspects should not correlate.

Overall, we believe that the testing and scoring methods for these studies should provide trustworthy data. The pilot study allowed refinement of testing measures, and rubrics were constructed and used for scoring free-response items. After confirming that our testing measures and scoring methods were reliable, we could proceed to the data analysis phase of this project in good faith.

3.6 Data Analysis Methods

After data collected, the data were analyzed using quantitative statistical methods. Study 1 data were collected in the fall semester of 2006. Study 2 data were collected in the spring semester of 2007. For each measure, the scores for each item were added to give a total score for each participant, and statistical measurements were performed on these score totals. Using score totals maximizes the variance for a given measure, but note that this can only be done for data that contain no missing values. All statistical analyses were performed using the SPSS statistical package. Descriptive statistics were used to obtain a general overview of the data and to see if the data had normal distributions. By using t-tests, we determined if scores of the experimental and control groups were significantly different for each construct. Correlation tests were used to find any significant relationships among the variables.
tested. Since a pretest-posttest design\textsuperscript{62} was used for most measures (excluding problem solving and confidence for class skills), a dependent $t$-test was used to determine if student’s skills/confidence were improving between testing periods.

### 3.6.1 Descriptive Statistics and Normality

Before performing the specific statistical tests required to address a research question, it is necessary to explore each data set as a whole. Descriptive statistics such as the mean and median describe the central tendency of the data. Standard deviation and range are two methods of reporting the spread of scores within a data set. Descriptive statistics allow for quick comparisons between data sets.

In addition, we can use the distribution of student scores to choose the most appropriate statistical tests for the data obtained. For example, parametric measures are built upon the assumption of a normal distribution.\textsuperscript{61} Some measures (e.g. $t$-tests) are more robust towards violations of the normality assumption than are others, so it is important to consider these factors before choosing a statistical test. When the assumptions of a parametric statistical test cannot be met, the researcher may need to use the equivalent nonparametric calculation. However, these are weaker and thus parametric measures should be used whenever possible.

The first step in determining if a data set has a normal distribution is to analyze the corresponding histogram. Skew and kurtosis are two descriptors that are commonly used to discuss distribution shapes, and these can be conveniently translated to numerical values. In order to visualize the relationship between these values and the distribution to which they correspond, a series of representative
distributions were generated and the skew and kurtosis were calculated for each one using SPSS (Figure 3.17). Skew is a measure of a distribution’s asymmetry. As shown below, a positively skewed distribution has a tail pointing towards higher values on the x-axis (points right). The tail for a negatively skewed distribution points towards lower values on the x-axis (points left). Kurtosis is defined as a distribution’s peakedness. If the data’s histogram is more peaked than a normal distribution, than it has a positive kurtosis value. If the data’s histogram is more flat than a normal distribution, then the data will have a negative kurtosis value.
Figure 3.17. Skew and Kurtosis for Various Distributions. To aid in visualizing skew and kurtosis, some distributions are shown below. Values of the skew and kurtosis for each distribution are given. Mean and standard deviation are also shown for comparison. Note that skew and kurtosis values are referenced to a normal distribution. A perfect normal distribution has a skew = 0.00 and a kurtosis = 0.00. Positive skew (Sk > 0) means a tail in the upper scores (points right), and negative skew (Sk < 0) means a tail in the lower scores (points left). Positive kurtosis (Kurt > 0) means a distribution more peaked than normal (with thinner tails), and negative kurtosis (Kurt < 0) means a distribution flatter than normal (with thicker tails).

\[n = 19\text{ for all distributions shown}\]
There are multiple ways to determine if a set of scores has a normal distribution. When looking at the skew and kurtosis values, it is important to note that the absolute value of the skew or kurtosis is less than two times its standard error in a normal distribution. This can be used as a general guideline, however this is not always the case for small data sets. Another indicator that can be used to find a normal distribution is a Q-Q (quantile-quantile) plot. This plots the quantiles of the given distribution against those of a normal distribution (or any distribution the user chooses). If the data set given has a normal distribution, then the points will lie close to the line (which lies along the diagonal of the plot). Kolmogorov-Smirnov tests can also be used to determine if the distribution in question significantly differs from a normal distribution. For this type of hypothesis test, we assume that the test distribution is the same as a standard normal distribution curve (a standard $H_0$ assumption). A value of $p > 0.05$ confirms that the distributions are the same (that the data are normal). Though there are multiple ways to determine normality, most investigators in educational and other research fields rely most heavily on the histogram itself (and the data’s skew and kurtosis statistics), and so the same protocol was followed for this assessment.

### 3.6.2 Independent Samples $t$-Tests

The $t$-test\textsuperscript{61} is used to determine if two distributions are different within the range of statistical error of the measurement. It is based on an assumption that there is no difference ($H_0$, the null hypothesis) between the two samples, and this must then be disproved within a 95% confidence level to give a significant $t$ statistic.
Significantly different distributions have $p$ values less than a predetermined alpha level. When $\alpha = 0.05$, the probability of incorrectly rejecting $H_0$ when it is actually true is 5 in 100. This is called type I error. There are several test statistics used to evaluate hypotheses about populations. The $t$-test statistic is based on a $t$-distribution (a specific normal distribution) and is used when the standard deviation for a population ($\sigma$) is unknown and therefore must be estimated from the sample data. The $t$-test is quite robust to deviations from normal and is suitable for even small sample sizes ($n < 25$).

A more conservative method of determining significance when using multiple statistical tests on the same data set is to use a corrected alpha level. The alpha level (usually $\alpha = 0.05$) set by the researcher for a statistical test represents the probability of a false positive, finding a difference between two groups when there is none. Since this is a probability, running multiple tests using the same alpha level would increase the likelihood of a false positive. The Bonferroni correction takes this into account and aims to minimize type I error. When a researcher runs five $t$-tests at $\alpha = 0.05$, she wants to have a 1 in 100 chance (0.05) of falsely rejecting the null hypothesis, not the 1 in five (0.22) chance that is actually expected due to random chance. Using the Bonferroni correction, the alpha level is corrected by dividing the desired alpha level ($\alpha = 0.05$) by the total number of statistical tests ($n_T$). For example, if we ran 5 $t$-tests ($n_T = 5$) on the same data set then the corrected alpha would be $0.05/5 = 0.01$. This keeps the alpha level of the overall experiment at a more acceptable level ($\alpha = 0.05$).

In this study, we used the same data sets to test not only for different
variables (depth of understanding, problem solving, etc.), but also to determine differences between an experimental and control group (independent $t$-tests) and differences between pretest and posttest scores (dependent $t$-tests). Thus the total number of statistical tests ($t$-tests) for each data set in our study was ten ($n_T = 10$), and so the corrected alpha of 0.005 was used for the Bonferroni corrected alpha level. Using this corrected alpha level, two means would be considered significantly different if $p < 0.005$ for a given $t$-test.

One drawback of using the Bonferroni correction is that while the probability of type I error (a false positive) is decreased, the probability of a type II error ($\beta$) is increased because the two types of error are related.\textsuperscript{74} Type II error is likelihood that $H_0$ will be accepted when it is actually false, meaning that the statistical test concludes that the groups are the same even though in reality the means of two groups are significantly different (a false negative). Anything that increases type II error, also compromises the power ($1-\beta$) for a study. For this and other reasons, some researchers argue against using corrections of alpha for multiple tests.\textsuperscript{75,74} In this work, both uncorrected and corrected data are reported for all $t$-tests.

### 3.6.3 Pearson’s Correlations

The hypothesis of this research was based on the expectation that certain variables were related. While it is impossible to determine causation, we can use correlation statistics to determine if two variables are linked or happen concurrently (high score for variable 1 happens at the same time as high score for variable 2). The most commonly used method of correlation is Pearson’s $r$ (also called Pearson’s
correlation coefficient). This can be used to detect linear correlations between two variables in either a positive or a negative direction. A positive value of $r$ means that as $x$ score increases, the value of $y$ also increases. There may also be an inverse relationship between $x$ and $y$, and this would give a negative $r$ value. For significant correlations ($p < 0.05$), the numerical value of Pearson’s $r$ is an indicator of the strength of an association. Cut-offs provided by Cohen\textsuperscript{76} can be used as a guideline for weak ($r > 0.1$), moderate ($r > 0.3$), and strong ($r > 0.5$) correlations. The value of $r_{xy}$ also relates to variance: $r^2$ is equal to the percent (\%) variance of the $y$ scores that is accounted for by a linear relationship with the $x$ scores.\textsuperscript{61}

### 3.6.4 Comparison of Pretest and Posttest Scores

In order to determine whether there has been an increase in students’ scores over time, it is necessary to compare scores at two different time periods. This is accomplished through a dependent \textit{t}-test (also called the paired-samples \textit{t}-test) on pretest and posttest scores for the corresponding measures. For this statistical measure, scores must be matched for each participant. Any participant that does not have scores for both pretest and posttest measures is excluded from this calculation. Pretest and posttest scores for each variable must be on the same number scale, otherwise converting to $z$-scores would be necessary. This statistical measure is similar to the one previously described in that it is used to determine if two data sets are significantly different ($p < 0.05$). For this NMR education assessment, pretest and posttest items were the same for the variables measured and used the same scale. However, some variables (problem solving and confidence class skills) did not
have pretest equivalents, and this was because students just entering the course were expected to have negligible scores in these areas. More than one exposure to the problem could itself lead to increased scores. This is called testing interference, and it is a validity threat that seems particularly likely for a difficult problem like the HETCOR problem solving task. Confidence for class skills was not tested because none of the students had performed any of the tasks listed, and similar information could be obtained from comparing the experimental (E) group that had completed the NMR module and control (C) group that had not.

3.6.5 Power Analysis

Power is important for t-test calculations and all other types of hypothesis testing in statistics. Statistical power is defined as the “probability of making a correct rejection” of the null hypothesis ($H_0$). This term is related to the type II error ($\beta$), the probability of retaining the $H_0$ hypothesis when it is false. When a type II error occurs the statistical test shows that two groups’ means are the statistically the same, when there is actually a difference. In any study, we would like to keep the power $(1 - \beta)$ high and the type I error ($\alpha$), the probability of rejecting $H_0$ when it is true, low, but there is often a trade-off between these two variables. One of the simplest ways to improve power while maintaining a sufficiently low probability of type I error ($\alpha \leq 0.05$) is to increase the number of participants ($n$).

Due to concerns brought about through the course of data analysis, researchers chose to run a post-hoc analysis of power on the data (section 4.3.2). For all power calculations, the G*Power statistical program was used. Power is
affected not only by the number of participants in a study (n), but also by the effect size of the variable measured. In the context of power analysis, effect size refers to the difference between population means (not the sample means) which is unknown and must be estimated. Cohen\textsuperscript{76} has reported a method of computing effect sizes (Cohen’s $d$), and recommendations for these effect sizes (t-tests: small 0.2, medium 0.5, large 0.8; correlations: small 0.1, medium 0.3, large 0.5). Power was computed for variables with small-medium effect sizes for t-tests and medium-large effect sizes for correlations (see section 4.3.2).
Chapter 4

Results and Discussion

4.1 Study 1: Fall 2006, Chem 457 and Chem 431

Data collected were analyzed according to the methods described in section 3.6, and the results were used to support or refute the hypothesis and to address the research questions proposed in this work.

4.1.1 Descriptive Statistics and Normality

Descriptive statistics for study 1 are shown below (Figure 4.1). For both cases (Chem 457 and Chem 431), the majority of means were centered near the middle of the range of given scores. However, problem solving scores are a notable exception. For this measure, most of the students had very low scores, most likely due to the difficulty of the HETCOR problem. Each course had a different set of class skills and so the items for these confidence measures (conf class skills) reflected this. For the conf class skills measure, the average score per item for Chem 457 students was 3.03. Students in Chem 431 rated their confidence slightly higher for their set of class skills with an average score of 4.34 per item.
**Figure 4.1. Descriptive Statistics for Study 1.** In this study there were two cases. Case 1 was Chem 457, the physical chemistry lab that completed the T1 experiment as their NMR module. Case 2 was Chem 431, the advanced organic and inorganic synthesis lab that completed the X-WIN tutorial as their NMR module. Descriptive statistics were used to obtain a general overview of the data and the distribution of scores for each data set.

**Case 1: Chem 457 F06 (n=19)**

*Posttest scores*

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Median</th>
<th>Range</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR Knowledge:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>4.74</td>
<td>1.41</td>
<td>5.00</td>
<td>2 - 8</td>
<td>0.657</td>
<td>1.03</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>11.68</td>
<td>2.93</td>
<td>12.0</td>
<td>6 - 16</td>
<td>-0.321</td>
<td>-0.973</td>
</tr>
</tbody>
</table>

| Problem-Solving:               |      |          |        |       |       |          |
| HETCOR Problem                 | 1.21 | 1.55     | 1.00   | 0 - 6 | 1.82  | 4.06     |

| Confidence:                    |      |          |        |       |       |          |
| Conf NMR tasks                 | 16.58| 7.47     | 19.0   | 5 - 30| -0.031| -0.949   |
| Conf Class Skills (NMR)        | 27.26| 15.0     | 29.0   | 9 - 58| 0.335 | -0.986   |
| Conf Gen Science Tasks         | 28.05| 9.73     | 31.0   | 6 - 39| -1.41 | 1.16     |

*Conf Class Skills (NMR)        | 27.26| 15.0     | 29.0   | 9 - 58| 0.335 | -0.986   |

| Conf Gen Science Tasks         | 28.05| 9.73     | 31.0   | 6 - 39| -1.41 | 1.16     |

|                      |      |          |        |       |       |          |
| std error            | 0.524|          |        |       |       |          |
| std error            | 1.01 |          |        |       |       |          |

**Case 2: Chem 431 F06 (n=24)**

*Posttest scores*

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Median</th>
<th>Range</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR Knowledge:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>4.42</td>
<td>1.72</td>
<td>4.50</td>
<td>1 - 7</td>
<td>-0.267</td>
<td>-0.162</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>13.71</td>
<td>2.90</td>
<td>13.0</td>
<td>7 - 19</td>
<td>-0.175</td>
<td>0.005</td>
</tr>
</tbody>
</table>

| Problem-Solving:               |      |          |        |       |       |          |
| HETCOR Problem                 | 2.92 | 1.64     | 2.00   | 1 - 6 | 0.920 | -0.498   |

| Confidence:                    |      |          |        |       |       |          |
| Conf NMR tasks                 | 19.92| 4.76     | 19.5   | 8 - 32| 0.235 | 1.91     |
| *Conf Class Skills (NMR)       | 43.75| 11.4     | 43.0   | 25 - 62| -0.013| 1.18     |
| Conf Gen Science Tasks         | 30.00| 4.17     | 29.0   | 21 - 38| 0.177 | 0.038    |

|                      |      |          |        |       |       |          |
| std error            | 0.472|          |        |       |       |          |
| std error            | 0.918|          |        |       |       |          |

* These conf class skills items were not the same set of items that was used for Chem 457 since class skills varied depending on the NMR module used.
It is interesting to note that Chem 431 students, as a group, had higher scores for both the problem solving task and basic knowledge of NMR. Responses from the pretest survey confirmed that most Chem 457 students in our study had already taken Chem 431 (and its prerequisites). However, students in Chem 431 have had more recent practice with basic NMR skills and spectral interpretation. Less time in between related activities may be one possible reason for the difference in basic NMR knowledge scores. Differences between the courses’ data sets may also arise from the participants themselves. For both data sets, the percent of students enrolled in the course who elected to participate in the study was approximately 50% (see section 3.3.1). With so many students not participating in the study, we considered the possibility of a self-selected sample. Self-selection is a type of selection bias that occurs when the group who choose to participate in the study differ in some way from eligible members of the target population. In study 1, extra credit was used as the only incentive, and students most likely chose to participate based on their perceived need for these extra points. However, it is possible that these two self-selected groups differed in what types of students felt they “needed” extra credit. Perhaps the Chem 457 participant group contained mainly low scoring students, those concerned worried about low grades in the course, while the Chem 431 participant group contained more students who were high scorers in their course and more motivated in general. Whatever the reasons for these differences, increasing the percentage of eligible students who participate in the studies would mitigate this concern.

For this NMR education project, we looked at a range of indicators to consider
the normality, but the final decision was always based on the histogram itself (for skew and kurtosis data, see Figure 4.1). Data for Chem 457 consisted of normal distributions for all variables except problem solving (HETCOR problem) and confidence for general science skills (conf gen science). For the HETCOR problem, zero was the most frequent score, but an extreme upper value (one score of 6) resulted in a mean greater than 1. The distribution for conf gen science was slightly negatively skewed, and it contained several extreme values on both ends. The data for Chem 431 differed in that confidence for NMR tasks (conf NMR) is the only variable with a distribution that deviates from normal. The histogram for this variable was more peaked in the center of the distribution than we would expect in a normal distribution. While there are some deviations from normal for the data in study 1, robust statistical measures (e.g. t-tests and Pearson's correlations) were still applicable for these data. There was no need to use nonparametric measures when analyzing the data for study 1.

4.1.2 Independent Samples t-Tests

Independent-samples t-test were used for each variable to determine if the mean of scores for students that had completed the NMR modules (the experimental group) was significantly greater than the mean of scores for those students who had not completed the module (the control group). The results of the t-test analyses for study 1 data are shown in Figure 4.2. Chem 431 students' confidence in class skills (NMR) was the only measure in this study shown to have significantly different means for the experimental (E) and control (C) groups $t(22) = 3.30, p < 0.01$. For
this measure, the mean of the E group \((m = 52.0, \, sd = 8.94)\) was significantly higher than that of the C group \((m = 38.8, \, sd = 9.81)\). The result for this variable and data set is also significant using Bonferroni's corrected alpha since \(p\) is less than the corrected alpha \((0.003 < 0.005)\). Since the control group had no exposure to the NMR class skills used throughout the NMR modules, this difference was an expected result. Researchers were hoping that this confidence might extend beyond the specific skills used into confidence for other NMR skills and perhaps further to confidence for general science skills. However, these measures did not yield significantly different results, and in fact showed differences in the opposite direction (mean of C group was greater than the mean of E) for Chem 457 students.
Figure 4.2. *t*-Tests for Equality of Means for Study 1 Data. The *t*-test for independent samples was used to determine if the mean of scores of the experimental group was significantly different from the mean scores of the corresponding control group for each data set. The results of the test are shown, as are the *t* value (with degrees of freedom), *p* value, and means (with standard deviation) of each group. Significantly different groups are denoted by *p* < 0.05.

**Case 1: Chem 457 F06 (n=19)**

*Posttest scores*

<table>
<thead>
<tr>
<th>Item Total</th>
<th><em>t</em>-Test: Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sig. <em>p</em> &lt; 0.05</td>
</tr>
<tr>
<td>NMR Knowledge:</td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>No</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>No</td>
</tr>
<tr>
<td>Problem Solving:</td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>No</td>
</tr>
<tr>
<td>Confidence:</td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>No</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>No</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>No</td>
</tr>
</tbody>
</table>

**Case 2: Chem 431 F06 (n=24)**

*Posttest scores*

<table>
<thead>
<tr>
<th>Item Total</th>
<th><em>t</em>-Test: Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sig. <em>p</em> &lt; 0.05</td>
</tr>
<tr>
<td>NMR Knowledge:</td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>No</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>No</td>
</tr>
<tr>
<td>Problem Solving:</td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>No</td>
</tr>
<tr>
<td>Confidence:</td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>No</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>Yes, E &gt; C</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>No</td>
</tr>
</tbody>
</table>

*m* = mean  
*sd* = standard deviation  
E = Experimental group  
C = Control group
For other measures essential to the research question (depth of understanding and problem solving), the independent-samples t-test did not yield significant results. However, the differences in the means were in the expected direction (mean of E > mean of C) for these variables. Since both cases in study 1 have relatively small numbers (n) of participants, it was possible that the lack of significance was due to low statistical power. Whether or not a lack of power was affecting study 1 results was unclear at this point, but the issue of power will be discussed further in section 4.3.2. It was also worthwhile to note that the two cases from study 1 (Chem 457 and Chem 431) yielded conflicting results for two variables, conf NMR skills and conf gen science skills. For Chem 457, the control group mean was higher than the experimental group mean for both variables (mean of C > mean of E), but Chem 431 results showed the opposite trends (mean of E > mean of C). A self-selected sample of low scorers in Chem 457 might explain this difference between the two cases (Chem 457 and Chem 431). Regardless, a second study with a broader coverage of eligible participants and more participants (higher n and greater power) would help to clarify these issues.

4.1.3 Pearson’s Correlations

Pearson’s correlation coefficients were calculated for the variables in study 1 to determine which variables might be related. The results of case 1 (Chem 457) are shown in Figure 4.3. For these data, basic knowledge of NMR concepts has been shown to correlate to both problem solving (strong positive correlation, \( r(17) = 0.580, p < 0.01 \)) and confidence for NMR tasks (moderate positive correlation, \( r(17) = 0.474, p < 0.01 \)).
In addition, there is a significant correlation (strong positive correlation, $r(17) = 0.573, p < 0.05$) between confidence for class skills (NMR) and confidence for NMR tasks not necessarily performed as part of the coursework.

**Figure 4.3. Pearson’s Correlation Coefficients for Study 1, Case 1 (Chem 457, Fall 2006).** The figure below shows the Pearson’s correlation coefficients (Pearson’s $r$) for all variable measures. Significant correlations are those with $p < 0.05$ and the level of significance is denoted by asterisks (*). The $r$ value shows the strength of the correlation, and this level is listed according to Cohen’s cut-offs (weak 0.1, moderate 0.3, strong 0.5).
<table>
<thead>
<tr>
<th>Case 1: Chem 457 F06</th>
<th>Depth of Understanding</th>
<th>Basic NMR Knowledge</th>
<th>HETCOR Problem Solving Task</th>
<th>Conf NMR Tasks</th>
<th>Conf Class Skills (NMR)</th>
<th>Conf Gen Science Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Understanding</td>
<td>$r = 1.00$</td>
<td>Not significant</td>
<td>Not significant</td>
<td>Not significant</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r (17) = 0.424$</td>
<td>$p = 0.071$</td>
<td>$r (17) = 0.205$</td>
<td>$p = 0.399$</td>
<td>$r (17) = 0.224$</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td></td>
<td>$r = 1.00$</td>
<td>strong</td>
<td>$r (17) = 0.580$</td>
<td>$p = 0.009^{**}$</td>
<td>$r (17) = 0.224$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r (17) = 0.474$</td>
<td>$p = 0.040^{*}$</td>
<td>$r (17) = 0.135$</td>
</tr>
<tr>
<td>HETCOR Problem Solving Task</td>
<td>$r = 1.00$</td>
<td></td>
<td></td>
<td>Not significant</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>Conf NMR Tasks</td>
<td></td>
<td></td>
<td></td>
<td>$r (17) = 0.162$</td>
<td>$p = 0.508$</td>
<td>$r (17) = 0.386$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r (17) = -0.007$</td>
<td>$p = 0.976$</td>
<td>$r (17) = -0.025$</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td></td>
<td></td>
<td></td>
<td>strong</td>
<td>$r (17) = 0.573$</td>
<td>$r (17) = -0.025$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r (17) = 0.010^{*}$</td>
<td>$p = 0.102$</td>
<td>$p = 0.920$</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td></td>
<td></td>
<td></td>
<td>$r = 1.00$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$N = 19$ for all variables
* Correlation significant at the 0.05 level (2-tailed)
** Correlation significant at the 0.01 level (2-tailed)
*** Correlation significant at the 0.001 level (2-tailed)
The few correlations found for Chem 457 (case 1) data were not what researchers were expecting. It was interesting to note that for this set of participants, knowledge of NMR concepts that were initially learned in textbooks (basic knowledge) was related to confidence for tasks that are largely experimental in nature (conf NMR tasks). Though researchers proposed that depth of understanding of NMR topics would affect problem solving (no significant correlation), a correlation between basic knowledge and problem solving was unexpected. Chem 457 (fall 2006) students had only very low scores for the HETCOR problem. While depth of understanding is expected to be crucial for a high score on the HETCOR problem, students that had a solid foundation in basic NMR knowledge (especially reading spectra skills) could be expected to generate responses that would obtain a non-zero problem solving score (scores of 1, 2 or 4 would be possible).

A correlation between confidence for class skills and confidence for NMR tasks was the only one that was expected and that seemed relevant to the hypothesis of this project. Researchers believed that hands-on NMR modules should lead to confidence in the skills used (class skills), and proposed that this confidence might be extended to more general NMR tasks. This may be the case as a positive correlation between both types of confidence was found. However, the link between the modules and this correlation remains unclear because there was not a significant difference (independent t-test) for the experimental and control groups for conf NMR skills for this group (see section 4.1.2).

As shown in Figure 4.4, Pearson’s correlation coefficients were also calculated
for case 2 (Chem 431) data and yielded several significant correlations. For this course, all confidence measures were positively correlated. Confidence for class skills (NMR) had significant correlations with both confidence for NMR tasks (strong positive correlation, $r(22) = 0.593, p < 0.01$) and confidence for general science tasks (strong positive correlation, $r(22) = 0.670, p < 0.001$). There was also a strong positive correlation between confidence for NMR tasks and confidence for general science tasks ($r(22) = 0.571, p < 0.01$).

Figure 4.4. Pearson’s Correlation Coefficients for Study 1, Case 2 (Chem 431, Fall 2006). The figure below shows the Pearson’s correlation coefficients (Pearson’s $r$) for all variable measures. Significant correlations are those with $p < 0.05$, and the level of significance is denoted by asterisks (*). The $r$ value shows the strength of the correlation, and this level is listed according to Cohen’s cut-offs (weak 0.1, moderate 0.3, strong 0.5).
<table>
<thead>
<tr>
<th></th>
<th>Depth of Understanding</th>
<th>Basic NMR Knowledge</th>
<th>HETCOR Problem Solving Task</th>
<th>Conf NMR Tasks</th>
<th>Conf Class Skills (NMR)</th>
<th>Conf Gen Science Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Understanding</td>
<td>$r = 1.00$</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td>Not Significant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r (22) = 0.393$</td>
<td>$r (22) = -0.33$</td>
<td>$r (22) = 0.222$</td>
<td>$r (22) = 0.278$</td>
<td>$r (22) = 0.273$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = 0.058$</td>
<td>$p = 0.877$</td>
<td>$p = 0.296$</td>
<td>$p = 0.189$</td>
<td>$p = 0.196$</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>$r = 1.00$</td>
<td></td>
<td>Not Significant</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td>Not Significant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$r (22) = 0.095$</td>
<td>$r (22) = -0.263$</td>
<td>$r (22) = 0.015$</td>
<td>$r (22) = 0.130$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p = 0.658$</td>
<td>$p = 0.214$</td>
<td>$p = 0.945$</td>
<td>$p = 0.546$</td>
</tr>
<tr>
<td>HETCOR Problem Solving</td>
<td>$r = 1.00$</td>
<td></td>
<td>Not Significant</td>
<td>Strong</td>
<td>Not Significant</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Task</td>
<td></td>
<td></td>
<td>$r (22) = -0.229$</td>
<td>Strong</td>
<td>$r (22) = 0.074$</td>
<td>$r (22) = -0.121$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p = 0.281$</td>
<td>Strong</td>
<td>$p = 0.733$</td>
<td>$p = 0.574$</td>
</tr>
<tr>
<td>Conf NMR Tasks</td>
<td></td>
<td></td>
<td></td>
<td>Strong</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r (22) = 0.593$</td>
<td>$p = 0.002**$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r (22) = 0.670$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$p = 0.000***$</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = 1.00$</td>
</tr>
</tbody>
</table>

N = 24 for all variables
* Correlation significant at the 0.05 level (2-tailed)
** Correlation significant at the 0.01 level (2-tailed)
*** Correlation significant at the 0.001 level (2-tailed)
As was found for the independent t-test data from study 1, the two courses gave opposite results for several Pearson’s r calculations. Variables that correlated positively with basic NMR knowledge for Chem 457 (case 1) showed weak correlations that were not significant for Chem 431 data. In fact, the correlation calculated to examine the relationship between basic NMR knowledge and confidence for NMR tasks resulted in a negative correlation that was not significant ($r(22) = -0.263, p = 0.214$). The Pearson’s r calculation for basic NMR knowledge and problem solving resulted in a negligible relationship that was not significant ($r(22) = 0.095, p = 0.658$).

Whether or not benefits (such as confidence) in one area of content can be extended to another is a sub-question of this research. The extension from confidence for (NMR) class skills specifically learned to those not experienced would be a positive result of this curriculum. The positive correlation between confidence in class skills and confidence for NMR tasks is consistent between the two courses. For Chem 431 students, the experimental group was shown to have a higher mean for class skills confidence (t-test) so it is possible that the modules are affecting not only class skills confidence, but also indirectly affecting confidence for NMR tasks. For this group, confidence for general science tasks was also shown to have a linear relationship with both confidence in NMR tasks and class skills (NMR). However, we cannot be sure whether one confidence is impacting another or whether some outside factor (perhaps specific to this course) is influencing both.
4.1.4 Comparison of Pretest and Posttest Scores

The results from the dependent $t$-tests calculated are shown in Figure 4.5. For these data, a positive $t$ value indicates that the mean posttest score was higher than the mean pretest score. This occurs when there is an increase in the scores over time. As can be seen, there were no significant differences in pretest and posttest scores for study 1. However, it should be noted that most of the differences in the means are in the expected direction (positive $t$). Since none of the dependent $t$-tests run for these data sets were found to be significant at the uncorrected alpha ($\alpha = 0.05$), there was no need to report data using the more conservative Bonferroni correction ($\alpha = 0.005$).
**Figure 4.5. Comparing Pretest and Posttest Measures (Study 1).** Dependent samples $t$-tests were used to compare pretest and posttest scores. Only matched scores were considered, meaning that only the scores from participants who participated in both surveys were used. The results of the test are shown, as are the $t$ value (with degrees of freedom), $p$ value, and means (with standard deviation) of each group. A positive value of $t$ indicates that the mean of the posttest score was higher than that of the pretest score. This would mean that scores increased over the course of the treatment. Significantly different groups would be indicated by $p < 0.05$.

**Case 1: Chem 457 F06 (n=18)*

*Matched Pretest and Posttest Scores*

<table>
<thead>
<tr>
<th>Item</th>
<th>Total</th>
<th>NMR Knowledge:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Depth of Understanding</td>
<td>$t(17) = 4.39$ (1.58)</td>
<td>$4.89$ (1.28)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Basic NMR Knowledge</td>
<td>$t(17) = 11.0$ (3.41)</td>
<td>$11.56$ (2.96)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem-Solving:</th>
<th>HETCOR Problem</th>
<th>N/A</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Confidence:</th>
<th>Conf NMR tasks</th>
<th>No $t = -0.064$</th>
<th>$0.950$</th>
<th>$16.5$ (7.05)</th>
<th>$16.4$ (7.64)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conf Class Skills (NMR)</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conf Gen Science Tasks</td>
<td>No $t = -0.678$</td>
<td>$0.507$</td>
<td>$29.7$ (5.04)</td>
<td>$28.3$ (9.93)</td>
</tr>
</tbody>
</table>

**Case 2: Chem 431 F06 (n=24)*

*Matched Pretest and Posttest Scores*

<table>
<thead>
<tr>
<th>Item</th>
<th>Total</th>
<th>NMR Knowledge:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Depth of Understanding</td>
<td>$t(23) = 4.33$ (2.01)</td>
<td>$4.42$ (1.72)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Basic NMR Knowledge</td>
<td>$t(23) = 13.5$ (3.09)</td>
<td>$13.7$ (2.90)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem-Solving:</th>
<th>HETCOR Problem</th>
<th>N/A</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Confidence:</th>
<th>Conf NMR tasks</th>
<th>No $t = 0.679$</th>
<th>$0.504$</th>
<th>$19.3$ (6.08)</th>
<th>$19.9$ (4.76)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conf Class Skills (NMR)</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conf Gen Science Tasks</td>
<td>No $t = 1.536$</td>
<td>$0.138$</td>
<td>$29.0$ (4.41)</td>
<td>$30.0$ (4.17)</td>
</tr>
</tbody>
</table>

*n = number of matched cases

N/A indicates that no pretest data were collected for this variable

$m = \text{mean} \quad \quad \quad sd = \text{standard deviation}$

Pre = Pretest scores \quad \quad Post = Posttest scores
There are two exceptions for which the mean of the pretest is higher than that of the posttest, and these are for case 1 (Chem 457) confidence for NMR tasks and general science skills. These are the same two variables that showed a reverse trend (C > E) with the independent samples t-test between the experimental and control groups. This suggests that confidence for NMR tasks may be decreasing as the course (Chem 457) progresses, and perhaps due to the NMR module itself. However, since these are not significant differences (for either t-test), researchers were hesitant to base conclusions on these data. Only if there is a significant difference and the same trends are seen in study 2 (with larger n) would this theory be considered plausible.

4.2 Study 2: Spring 2007, Chem 457

4.2.1 Descriptive Statistics and Normality

Figure 4.6 shows the descriptive statistics for study 2 (Chem 457, Spring 2007) and the statistics for the same course in the fall from study 1 (case 1). The primary difference between these two data sets is the number of participants. The instructor and course work remained the same for both semesters, and yet some noticeable differences in the scores were found. Overall, the mean scores from study 2 were higher than the mean scores from study 1 for all variables. The smaller percent of participating students in the fall classes (< 50% of enrolled) from study 1 may represent a self-selected sample (see section 4.1.1). In this case (study 1), the concern was that students who were enticed by extra credit alone were not representative of the class (Chem 457 or Chem 431) as a whole. In study 2, 83% of
students enrolled in Chem 457 in the spring of 2007 (study 2) completed the posttest (93% participated in this study overall) in part due to added incentives (iPod drawing). Similar to the fall data, problem solving for study 2 was the only variable for which the mean was not centered near the middle of the range. This confirmed that the HETCOR task is extremely difficult for all students.
Figure 4.6. Descriptive Statistics for Study 2 (Chem 457). In this study there was only one case, Chem 457 (spring 2007), the physical chemistry lab that completed the T1 experiment as their NMR module. Descriptive statistics were used to obtain a general overview of the data and the distribution of scores for each data set. Since this is the same course used in study 1 (case 1), the statistics from that data set are also shown to facilitate comparisons. More details on study 1 data can be found in section 4.1.1.

Study 2:
Chem 457 S07 (n=70)
Posttest scores

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Median</th>
<th>Range</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR Knowledge:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>5.00</td>
<td>1.55</td>
<td>5.00</td>
<td>1 - 8</td>
<td>-0.264</td>
<td>-0.092</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>12.40</td>
<td>3.63</td>
<td>12.0</td>
<td>4 - 19</td>
<td>-0.099</td>
<td>-0.821</td>
</tr>
<tr>
<td>Problem Solving:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>2.42</td>
<td>1.83</td>
<td>2.00</td>
<td>0 - 7</td>
<td>0.705</td>
<td>-0.287</td>
</tr>
<tr>
<td>Confidence:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>18.43</td>
<td>6.26</td>
<td>19.0</td>
<td>5 - 30</td>
<td>-0.403</td>
<td>-0.478</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>34.89</td>
<td>13.2</td>
<td>36.0</td>
<td>9 - 56</td>
<td>-0.342</td>
<td>-0.843</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>30.67</td>
<td>5.82</td>
<td>32.0</td>
<td>6 - 41</td>
<td>-1.63</td>
<td>4.59</td>
</tr>
</tbody>
</table>

Study 1, Case 1:
Chem 457 F06 (n=19)
Posttest scores

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Median</th>
<th>Range</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR Knowledge:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>4.74</td>
<td>1.41</td>
<td>5.00</td>
<td>2 - 8</td>
<td>0.657</td>
<td>1.03</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>11.68</td>
<td>2.93</td>
<td>12.0</td>
<td>6 - 16</td>
<td>-0.321</td>
<td>-0.973</td>
</tr>
<tr>
<td>Problem Solving:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>1.21</td>
<td>1.55</td>
<td>1.00</td>
<td>0 - 6</td>
<td>1.82</td>
<td>4.06</td>
</tr>
<tr>
<td>Confidence:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>16.58</td>
<td>7.47</td>
<td>19.0</td>
<td>5 - 30</td>
<td>-0.031</td>
<td>-0.949</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>27.26</td>
<td>15.0</td>
<td>29.0</td>
<td>9 - 58</td>
<td>0.335</td>
<td>-0.986</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>28.05</td>
<td>9.73</td>
<td>31.0</td>
<td>6 - 39</td>
<td>-1.41</td>
<td>1.16</td>
</tr>
</tbody>
</table>

std error: 0.287
std error: 0.566

std error: 0.524
std error: 1.01
When considering the normality of study 2 data, researchers relied most heavily on the histogram itself (and the data’s skew and kurtosis statistics) to decide which statistical tests should be used for the data analysis. There were only two variables for which the normality of the distribution was questionable in study 2. These were the same variables found to deviate from normal in the first study (for case 1): problem solving (HETCOR problem) and confidence for general science skills (conf gen science). The distribution for the problem solving task was bimodal (with modes at 1 and 2), and was slightly positively skewed. The data for confidence for general science skills had a high kurtosis value, but this high value most likely arises from two extreme low scores. These were quite far from the bulk of responses and resulted in a long tail on the left end of the histogram. Despite these outliers, these deviations from normal were still reasonable since robust tests such as the \( t \)-test and Pearson’s \( r \) correlation were to be used.

4.2.2 Independent Samples \( t \)-Tests

To determine if there were significant differences between the mean scores for the experimental and control groups, independent samples \( t \)-tests were calculated for study 2 data (Figure 4.7). For this data set, the experimental means were significantly higher than the control means for three variables. Depth of understanding scores were shown to be significantly different for the two groups (\( t(68) = 2.15, p < 0.05 \)) with the mean of the experimental group (\( m = 5.29, sd = 1.58 \)) being greater than that of the control group (\( m = 4.48, sd = 1.39 \)). There was also a significant difference between the groups for confidence for NMR tasks (\( t(68) = 2.33, \))
The mean of experimental group’s scores ($m = 19.7$, $sd = 5.92$) were higher than the mean of the control group’s scores ($m = 16.2$, $sd = 6.32$). Similar to study 1 results, the experimental group had significantly ($t(68) = 7.08$, $p < 0.001$) higher scores than the control group for confidence for class skills. This variable had the largest difference in means found (E group: $m = 41.3$, $sd = 10.3$; C group: $m = 23.4$, $sd = 9.88$).

**Figure 4.7. t-Tests for Equality of Means for Study 2 Data (Chem 457, Spring 2007).** The $t$-test for independent samples was used to determine if the mean of the experimental group was significantly different from the mean of the corresponding control group for each data set. The results of the tests are shown. For each measure, the $t$ value (with degree of freedom), $p$ value, and means (with standard deviation) of each group are also given. Significantly different groups are denoted by $p < 0.05$.

**Chem 457 S07 (n=70)**  
Posttest scores

<table>
<thead>
<tr>
<th>Item Total</th>
<th>$t$-Test: Equality of Means</th>
<th>$t$ (68)</th>
<th>$p$</th>
<th>$m(sd)$ of E</th>
<th>$m(sd)$ of C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NMR Knowledge:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>Yes, E &gt; C</td>
<td>2.15</td>
<td>0.036</td>
<td>5.29 (1.58)</td>
<td>4.48 (1.39)</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>No</td>
<td>1.03</td>
<td>0.305</td>
<td>12.7 (3.73)</td>
<td>11.8 (3.43)</td>
</tr>
<tr>
<td><strong>Problem-Solving:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>No</td>
<td>1.03</td>
<td>0.305</td>
<td>2.56 (1.63)</td>
<td>2.08 (2.18)</td>
</tr>
<tr>
<td><strong>Confidence:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>Yes, E &gt; C</td>
<td>2.33</td>
<td>0.023</td>
<td>19.7 (5.92)</td>
<td>16.2 (6.32)</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>Yes, E &gt; C</td>
<td>7.08</td>
<td>0.000</td>
<td>41.3 (10.3)</td>
<td>23.4 (9.88)</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>No</td>
<td>0.588</td>
<td>0.558</td>
<td>31.0 (5.39)</td>
<td>30.1 (6.60)</td>
</tr>
</tbody>
</table>

$m =$ mean  
$sd =$ standard deviation  
E = Experimental group  
C = Control group

Some results from study 2 differ from those seen for the smaller subset of students in the fall Chem 457 course (study 1, case 1). This gives more evidence to
the theory that the students from study 1 (case 1) were a self-selected sample who most likely differed from the general population of Penn State students (who would take Chem 457). The most obvious difference between the studies was that there were no negative $t$ values for study 2, meaning that the experimental group scored higher on average than the control group for all measures. In the case of confidence for NMR tasks, not only was the difference in the positive ($E > C$) direction for study 2, but it was also significant. Other measures were significantly different for the larger group (study 2) but not significant for the smaller group (study 1). Different results for the same course (Chem 457) between study 1 and 2 could suggest a lack of power for the first sets of data (study 1) for these cases (depth of understanding, confidence for NMR tasks, and confidence class skills), and will be discussed further in section 4.3.2. For those variables that were not found to be significantly different in study 2 (e.g. problem solving), it is unclear whether there is truly no difference or whether the effect size is just too small to be measured.

The results from these independent $t$-tests were very promising. From these we can conclude that the NMR modules (at least for Chem 457) are working as designed as increase student’s depth of understanding for NMR. It was proposed that the modules would have an effect on students’ problem solving skills (according to the hypothesis of this work), but no significant difference was seen for this measure. The modules also have a positive effect on confidence for class skills, as would be expected since students in the experimental group were able to practice these skills. Those designing the NMR modules believed that they would improve student confidence for not only class skills, but also confidence for other NMR tasks
and perhaps even confidence for general science skills. There was a positive effect found for the former (NMR tasks) that represents an extension of confidence beyond what would be expected. However, it seems that confidence for general science skills was not affected by interaction with the NMR modules. Since the sample consists of upper level science majors, it may be that their confidence in these skills is already relatively stable. This may be one reason that we do not see an effect on this variable.

Using the more conservative cut-off for type I error (Bonferroni’s correction) revealed different results for study 2. A comparison of the results for independent samples t-test on these data found with and without using the Bonferroni correction is shown in Figure 4.8. While the result for confidence for class skills remains the same, the E and C groups were not found to be significantly different for either depth of understanding and confidence for NMR tasks using the Bonferroni corrected alpha level. This suggests that the uncorrected results for the latter two variables may be false positives. On the other hand, using this correction is known to affect power and the probability of a type II error (a false negative). It should be noted that all conclusions presented in this work were based on the uncorrected data.
Figure 4.8. Bonferroni Correction for t-Test Equality of Means Data (Study 2). Using the p values calculated from the independent t-tests shown above, results using the corrected α were determined. Results determined using the uncorrected α are also shown to enable comparisons. The Bonferroni corrected alpha* was calculated as shown.

### Chem 457 S07 (n=70)
Posttest scores

<table>
<thead>
<tr>
<th>Item</th>
<th>Total</th>
<th>t-Test: Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Significant using α</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sig. p &lt; 0.05</td>
</tr>
<tr>
<td>NMR Knowledge:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td></td>
<td>Yes, E &gt; C</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Problem-Solving:</td>
<td></td>
<td>HETCOR Problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Confidence:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td></td>
<td>Yes, E &gt; C</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td></td>
<td>Yes, E &gt; C</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

α = 0.05 *Corrected α = 0.005

*Bonferroni Correction
Corrected α = α/n<sub>T</sub>

n<sub>T</sub> = 10     0.05/10 = 0.005

### 4.2.3 Pearson’s Correlations

To determine if there were statistically significant linear relationships between any of the variables for study 2, Pearson’s correlation coefficients were calculated (Figure 4.9). Significant positive correlations were found for most bivariate calculations among variables for this data set. There was a positive correlation (moderate correlation, r(68) = 0.319, p < 0.01) between depth of
understanding and problem solving. Problem solving was also found to correlate with confidence for NMR tasks (moderate correlation, $r(68) = 0.316, p < 0.01$) and confidence for class skills (weak correlation, $r(68) = 0.280, p < 0.05$), but not with confidence for general science skills (correlation not significant, $p > 0.05$). As was found for previous data sets (study 1: case 1 and 2), there was a strong correlation ($r(68) = 0.614, p < 0.01$) between confidence for class skills (NMR) and confidence for more general NMR tasks.

Figure 4.9. Pearson’s Correlation Coefficients for Study 2 (Chem 457, Spring 2007). The figure below shows the Pearson’s correlation coefficients (Pearson’s $r$) for all variables. Significant correlations are those with $p < 0.05$ and the levels of significance are denoted by asterisks (*). The $r$ value shows the strength of the correlation, and this level is listed according to Cohen’s cut-offs (weak 0.1, moderate 0.3, strong 0.5).
### Study 2: Chem 457  
Spr 07
Depth of Understanding  
Basic NMR Knowledge  
HETCOR  
Problem-Solving Task  
Conf NMR Tasks  
Conf Class Skills (NMR)  
Conf Gen Science Tasks

<table>
<thead>
<tr>
<th></th>
<th>Depth of Understanding</th>
<th>Basic NMR Knowledge</th>
<th>HETCOR Problem-Solving Task</th>
<th>Conf NMR Tasks</th>
<th>Conf Class Skills (NMR)</th>
<th>Conf Gen Science Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Understanding</td>
<td>$r = 1.00$</td>
<td>moderate</td>
<td>moderate</td>
<td>moderate</td>
<td>weak</td>
<td>weak</td>
</tr>
<tr>
<td></td>
<td>$r (68) = 0.412$</td>
<td>$r (68) = 0.319$</td>
<td>$r (68) = 0.322$</td>
<td>$r (68) = 0.351$</td>
<td>$r (68) = 0.252$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.000^{**}$</td>
<td>$p = 0.007^{**}$</td>
<td>$p = 0.006^{**}$</td>
<td>$p = 0.003^{**}$</td>
<td>$p = 0.035^{*}$</td>
<td></td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = 1.00$</td>
<td>moderate</td>
<td>weak</td>
<td>strong</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r (68) = 0.412$</td>
<td>$r (68) = 0.362$</td>
<td>$r (68) = 0.280$</td>
<td>$r (68) = 0.402$</td>
<td>Not significant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.000^{***}$</td>
<td>$p = 0.002^{**}$</td>
<td>$p = 0.019^{*}$</td>
<td>$p = 0.052$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem-Solving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>$r = 1.00$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf NMR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 70 for all variables  
* Correlation significant at the 0.05 level (2-tailed)  
** Correlation significant at the 0.01 level (2-tailed)  
 *** Correlation significant at the 0.001 level (2-tailed)
The significant Pearson’s correlation coefficient for study 2 have interesting implications for this NMR assessment. The hypothesis of this work was that depth of understanding in NMR might lead to increased problem solving skills. There is indeed a correlation between these measures, however it is not a strong correlation ($r(68) = 0.319$). Even if depth of understanding has an effect on students’ problem solving abilities for the HETCOR task, it could only be a small contributor. Small effect size might explain why there was not a measurable difference in the experimental and control groups scores (independent $t$-test, section 4.2.2) for problem solving even though there was a significant difference for depth of understanding (E > C).

Through the hypothesis of this NMR assessment, it was proposed that confidence might be an intermediary between depth of understanding (of NMR) and increased problem solving. Indeed, there is evidence that confidence is related to both variables. All three types of confidence were found to correlate to the depth of understanding measure. In addition, moderate-weak correlations were found between problem solving and both confidence for NMR tasks and confidence for class skills (NMR). There was not a significant Pearson’s correlation between problem solving and confidence for general science skills, and this may have been expected since the set of general science skills did not specifically include problem solving skills.

All three types of confidence measured were found to have significant correlations. The strongest correlation was between confidence for class skills (NMR) and confidence for NMR tasks, and this result was consistent across all data sets.
(study 1 and 2). Students’ self-reported confidence for tasks they had not completed (conf NMR tasks) was affected by their confidence for class skills, and this effect represents a meaningful product of these NMR modules. It was not surprising that the correlation was weaker between confidence for class skills and confidence for general science tasks, as the NMR modules do not seem to have a direct effect on the latter (independent t-tests, section 4.2.2).

4.2.4 Comparison of Pretest and Posttest Scores

Comparisons of pretest and posttest scores using dependent t-tests were performed to determine if the scores of students in study 2 (Chem 457, spring 2007) were improving over time. The results of these tests are shown below (Figure 4.10). For this data set, all of the t values were positive which indicates that mean posttest scores were higher than mean pretest scores for all variables. There was a significant increase ($t(67) = 3.367, p < 0.01$) in scores for depth of understanding in study 2. The mean of the posttest measure was 5.00 ($sd = 1.57$), and the mean of the pretest measure was 4.26 ($sd = 1.45$). Participants in study 2 also had significantly higher ($t(67) = 2.247, p < 0.05$) posttest scores for the confidence for the NMR tasks measure (pretest: $m = 18.3, sd = 6.19$; posttest: $m = 16.6, sd = 6.18$).
Figure 4.10. Comparing Pretest and Posttest Measures (Study 2).

Dependent samples \( t \)-tests were used to compare pretest and posttest scores for relevant variables. Only matched scores were considered, meaning that only the scores from students who participated in both surveys were used. The results of the \( t \)-tests are shown. Also, the \( t \) value (with degree of freedom), \( p \) value, and means (with standard deviation) of each group are given for each measure. A positive value of \( t \) indicates that the mean of the posttest score was higher than that of the pretest score, and confirms that scores increased over the course of the treatment. Significantly different groups are indicated by \( p < 0.05 \).

Chem 457 S07  \( (n=68)* \)

Matched Pretest and Posttest Scores

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Dependent ( t )-Test: Paired Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \text{sig. } p &lt; 0.05 )</td>
</tr>
<tr>
<td>NMR Knowledge</td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>Yes, increase</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>NO</td>
</tr>
<tr>
<td>Problem-Solving:</td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>N/A</td>
</tr>
<tr>
<td>Confidence</td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>Yes, increase</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>N/A</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>NO</td>
</tr>
</tbody>
</table>

\*n = number of matched cases
N/A indicates that no pretest data were collected for this variable

\( m \) = mean  \( sd \) = standard deviation
Pre = Pretest scores  Post = Posttest scores

These dependent \( t \)-test calculations have positive implications for the NMR project as implemented. Students’ depth of understanding of NMR concepts and confidence for NMR tasks were both positively effected by the new hands-on modules. Not only can we say that the experimental group had better scores for these measures, but also we now know that students’ scores improved over the testing period for both variables. A significant difference between these means for the variable gives evidence that researchers have successfully designed NMR.
modules that increase students’ depth of understanding of NMR. Also there is evidence supporting the sub-hypothesis that increased depth of understanding is linked to an increase in students’ confidence for NMR tasks.

Factors that did not change over time were basic NMR knowledge and confidence for general science tasks. The data suggested that interaction with the NMR modules also had no effect on these variables (independent t-test calculations for the experimental and control group, section 4.2.2). It seems these variables are quite stable for these samples (upper level science majors), and cannot be altered significantly through one NMR module.

The Bonferroni correction was used on the dependent t-tests calculated above, and these results are shown in Figure 4.11. For this set of tests, depth of understanding was found to be significant using the corrected alpha (α= 0.005), meaning there was a significant increase in scores from the pretest to the posttest. However, confidence for NMR tasks was not found to be significantly different for the pretest and posttest scores when using Bonferroni’s correction. Considering that this is a more conservative method, this result is not surprising.
Figure 4.11. Bonferroni Correction for Comparing Pretest and Posttest Data (Study 2). Using the \( p \) values calculated from dependent \( t \)-tests shown above, results using the corrected \( \alpha \) were determined. Results determined using the uncorrected \( \alpha \) are also shown to enable comparisons. The Bonferroni corrected alpha* was calculated as shown.

<table>
<thead>
<tr>
<th>Item Total</th>
<th>Dependent ( t )-Test: Paired Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Significant using ( \alpha )</td>
</tr>
<tr>
<td>NMR Knowledge:</td>
<td></td>
</tr>
<tr>
<td>Depth of Understanding</td>
<td>Yes, increase</td>
</tr>
<tr>
<td>Basic NMR Knowledge</td>
<td>NO</td>
</tr>
<tr>
<td>Problem Solving:</td>
<td></td>
</tr>
<tr>
<td>HETCOR Problem</td>
<td>N/A</td>
</tr>
<tr>
<td>Confidence:</td>
<td></td>
</tr>
<tr>
<td>Conf NMR tasks</td>
<td>Yes, increase</td>
</tr>
<tr>
<td>Conf Class Skills (NMR)</td>
<td>N/A</td>
</tr>
<tr>
<td>Conf Gen Science Tasks</td>
<td>NO</td>
</tr>
</tbody>
</table>

*Bonferroni Correction \( \alpha = 0.05 \) \( \alpha = \frac{0.05}{10} = 0.005 \)

4.3 Summary of Results and Further Discussion

4.3.1 Data Summary

A summary of the data collected for the NMR curriculum assessment is shown in Figure 4.12 below. The data are arranged by course to enable comparisons between the two studies. Before discussing how these findings relate to the research questions, a brief discussion of the differences among the studies is warranted. Two
implications drawn from these differences are:

1. Researchers should be wary of self-selected samples when the ratio/percent of participants to students enrolled in the course is small. The participants in study 1 (case 1) may be a self-selected sample, as seen by difference between these Chem 457 student scores and those found for study 2. In addition, unexpected trends when comparing data from Chem 457 (study 1, case 1) and data from Chem 431 (study 1, case 2) were found that may also imply a self-selected sample for study 1 (case 2), considering that Chem 431 was taken by most Chem 457 participants earlier in their undergraduate careers.

2. The methods used for study 2 should serve as a model for studies of this type. Not only did this study have more significant calculations, it also seemed to avoid the inconsistent results often found for self-selected samples.

**Figure 4.12. Data Summary for NMR Assessment Project (Studies 1 & 2).** A summary of the data collected for this work is shown below. The mean scores for each data set are given (with standard deviation in italics). The independent *t*-tests show whether the experimental (E) or control (C) group had the higher mean, and also whether the results were significant (SIG). Dependent *t*-tests were used to compare pretest (Pre) and posttest (Post) mean scores. Only matched cases were used for these calculations and *n*(matched) is given for each data set (denoted by an asterisk*). The relationship among the mean scores for Pre and Post is shown, and significant *t*-tests are indicated (SIG). For the Pearson’s correlation coefficient calculations (Pearson’s *r*), only significant correlations are shown. Variables are numbered (1-6) on the left of the chart, and these can be matched to see which bivariate correlations were significant. Correlations are arranged by strength of correlation (strong S, moderate M, weak W) according to Cohen’s recommended cut-offs.
Chem 457
Study 1 (Case 1), Chem 457 Fall 2006 (n = 19)
Study 2, Chem 457 Spring 2007 (n = 70)

NMR Knowledge:

<table>
<thead>
<tr>
<th>Study</th>
<th>NMR Knowledge</th>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
<td>E vs. C</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>1. Depth of Understanding</td>
<td>4.74</td>
<td>5.00</td>
<td>E &gt; C</td>
<td>NS</td>
<td>E &gt; C</td>
</tr>
<tr>
<td>2. Basic NMR Knowledge</td>
<td>11.68</td>
<td>12.40</td>
<td>E &gt; C</td>
<td>E &gt; C</td>
<td>Post &gt; Pre</td>
</tr>
</tbody>
</table>

Problem-Solving:

3. HETCOR Problem

<table>
<thead>
<tr>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>1.21</td>
<td>2.42</td>
<td>E &gt; C</td>
<td>E &gt; C</td>
</tr>
</tbody>
</table>

Confidence:

4. Conf NMR tasks

<table>
<thead>
<tr>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>16.58</td>
<td>18.43</td>
<td>E &gt; C</td>
<td>SIG</td>
</tr>
<tr>
<td>27.26</td>
<td>34.89</td>
<td>E &gt; C</td>
<td>SIG</td>
</tr>
</tbody>
</table>

Conf Gen Science Tasks

<table>
<thead>
<tr>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>28.05</td>
<td>30.67</td>
<td>C &gt; E</td>
<td>E &gt; C</td>
</tr>
</tbody>
</table>

Chem 431
Study 1 (Case 2), Chem 431 Fall 2006 (n = 24)

NMR Knowledge:

<table>
<thead>
<tr>
<th>Study</th>
<th>NMR Knowledge</th>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
<td>E vs. C</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>4.42</td>
<td>1.72</td>
<td>E &gt; C</td>
<td>NS</td>
<td>Post &gt; Pre</td>
<td>NS</td>
</tr>
</tbody>
</table>

Problem-Solving:

3. HETCOR Problem

<table>
<thead>
<tr>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>2.92</td>
<td>1.64</td>
<td>E &gt; C</td>
<td>NS</td>
</tr>
</tbody>
</table>

Confidence:

4. Conf NMR tasks

<table>
<thead>
<tr>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>19.92</td>
<td>4.76</td>
<td>E &gt; C</td>
<td>Post &gt; Pre</td>
</tr>
<tr>
<td>43.75</td>
<td>11.4</td>
<td>E &gt; C</td>
<td>SIG</td>
</tr>
</tbody>
</table>

Conf Gen Science Tasks

<table>
<thead>
<tr>
<th>Mean</th>
<th>Independent t-Tests*</th>
<th>Dependent t-Tests*</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation</td>
<td>E vs. C</td>
<td>significant?</td>
</tr>
<tr>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>30.00</td>
<td>4.17</td>
<td>E &gt; C</td>
<td>SIG</td>
</tr>
</tbody>
</table>

NS = No significant difference found
SIG = Significant difference found
*Study 1: n (matched) = 17
*Study 2: n (matched) = 68
** This is not the same set of items as was used for Chem 457 since class skills varied depending on the NMR module used.

E = Experimental group
C = Control group
Pre = Pretest scores
Post = Posttest scores
The students in Chem 457 (study 1, case 1 and study 2) had the most similar testing experience, and yet the results were slightly different for these two data sets. As was previously mentioned, we were concerned that participants in study 1 may have been a self-selected sample because a smaller percent of students in the fall Chem 457 classes participated (< 50% of enrolled) in the study (study 1, case 1). In contrast, 83% of students enrolled in Chem 457 in the spring of 2007 (study 2) completed the posttest (93% participated in this study overall). Many of these statistical calculations showed similar trends, but were not significant for study 1 (case 1) data. In addition, several variables, confidence for NMR tasks and confidence for general science tasks, have opposite trends among the two studies. This combined with the noticeably lower mean scores for all variable measures in study 1 (case 1) lead us to conclude that we indeed had a self-selected sample for this data set. We believe that if a broader sampling of fall semester students had participated that the trends would have been consistent between both studies.

The methods of study 2 were meant to improve upon those used in study 1. Chem 457 was chosen as the course to be modified for several reasons. For this course, NMR content came only from the researchers, and there were no other NMR experiences within the course that could interfere with testing. In Chem 431, students use the NMR spectrometer throughout the entire semester, and it is indispensable for monitoring reactions and identifying unknowns. This kind of overlap could lead to noticeable interference, and was the reason that online testing was carried out at the beginning of the semester for this course. In addition, the grant written for this project focused on changes to physical chemistry laboratories,
and so Chem 457 would be the most relevant to assessment required by the funding agency. Chem 457 has higher enrollment and is offered both semesters every year providing a greater opportunity for the NMR curriculum to affect students at PSU.

### 4.3.2 Power Analysis

Considering the low numbers of participants in study 1 and the lack of significant differences found, we were concerned that these cases (case 1 and 2) may have suffered from a lack of statistical power. However, to determine whether or not this was a reasonable concern, we conducted a power analysis$^{77,81}$ of the data using the methods described in section 3.6.5. The post-hoc analysis was completed for each data set for several effect sizes (Figure 4.13) using G*Power statistical program.$^{78,79}$ As previously mentioned, effect size in a power analysis relates to the difference between population means which is unknown and must be estimated. Using Cohen’s$^{76}$ recommendations for effect sizes (t-tests: small 0.2, medium 0.5, large 0.8; correlations: small 0.1, medium 0.3, large 0.5), power was computed for variables with small-medium effect sizes for t-tests and medium-large effect sizes for correlations.
**Figure 4.13. Post-hoc Analysis of Power (Studies 1 & 2).** The estimated power for studies 1 and 2 were calculated using the post-hoc method for computing achieved power in the G*Power statistical program.\textsuperscript{78,79} The power achieved was calculated for a two-tailed test with an alpha level of 0.05 ($\alpha = 0.05$) by entering the number of participants ($n$) for the corresponding data set. Values were calculated at two effect size levels (0.3 and 0.5) for all three data sets.

<table>
<thead>
<tr>
<th>Study</th>
<th>Case 1: Chem 457</th>
<th>Case 2: Chem 431</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$-Test:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Independent</td>
<td>Dependent</td>
</tr>
<tr>
<td></td>
<td>Groups</td>
<td>Groups</td>
</tr>
<tr>
<td>Study 1</td>
<td>0.177</td>
<td>0.541</td>
</tr>
<tr>
<td></td>
<td>0.206</td>
<td>0.650</td>
</tr>
<tr>
<td>Study 2</td>
<td>0.506</td>
<td>0.985</td>
</tr>
</tbody>
</table>

**Post-hoc: Power Achieved**

0.3 effect size*

<table>
<thead>
<tr>
<th>Study</th>
<th>Chem 457</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$-Test:</td>
</tr>
<tr>
<td></td>
<td>Independent</td>
</tr>
<tr>
<td></td>
<td>Groups</td>
</tr>
<tr>
<td></td>
<td>Dependent</td>
</tr>
<tr>
<td></td>
<td>Groups</td>
</tr>
<tr>
<td>Study 1</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>0.104</td>
</tr>
<tr>
<td>Study 2</td>
<td>0.220</td>
</tr>
</tbody>
</table>

* $\alpha = 0.05$, two-tailed test

As can be seen in Figure 4.13, power increases with the number of participants ($n$) and higher effect sizes. Study 1 had fewer participants for both cases (case 1, $n = 19$; case 2, $n = 24$), but participation in study 2 was noticeably greater ($n = 70$). A commonly accepted cut-off for power is 0.8,\textsuperscript{82} and while this was not achieved for any tests in study 1, several of the power values for study 2 were found to be above this cut-off. For these tests (dependent $t$-test and correlations), there was
sufficient power and a good chance of correctly rejecting the null hypothesis at least for any variable with an effect size of 0.5 or greater. However, it should be noted that with a smaller effect size (0.3), power was likely insufficient even for study 2. This means that variables with a low-medium effect size (t-test) may were more likely to go undetected in this study.

Post-hoc power analysis for independent t-tests revealed that power would be insufficient for the conditions calculated, and so further analysis was required. How many participants would be needed to achieve a suitable power level? Would 100% participation in smaller courses (such as Chem 431) still fail to give a power level of 0.8? Using G*Power, a plot was constructed of power versus sample size for several effect sizes (0.3 and 0.5) to answer these questions (Figure 4.14). Data suggest that to reach the desired 0.8 level of power, a study would need more than 100 participants to detect a variable that has a 0.5 effect size. While this would be possible for Chem 457 (spring semester has 120 openings), Chem 431 has a maximum enrollment of 75 per semester. Only if results from multiple semesters of the same course could be combined would this be achievable at Penn State. The number of participants needed to detect a variable with a 0.3 effect size is much higher (~ 350). It would take two to three years of collecting data (over multiple semesters) to reach this sample size.
Figure 4.14. Analysis of Power versus Sample Size for an Independent Samples t-Test. The general trend for power versus sample size for an independent samples t-test is shown below for several effect sizes. The data used to construct this graphs were generated using the G Power statistical program. Power was calculated for a two-tailed test with an alpha level of 0.05 ($\alpha = 0.05$). As can be seen from this graph, the sample size required for a given power level depends heavily of the expected effect size.

Despite some concerns, we believe that if the same methods as study 2 are used, similar results should be obtained for future studies and other courses (specifically Chem 431). Even though power was somewhat low for independent t-tests, the resultant power for study 2 seems reasonable given the constraints of this NMR curriculum study. Several variables were found to have significantly different means (independent t-tests) which is informative, even if those variables with lower effect sizes might not be detected. A greater number of participants in study 2 led to higher power, and this increase in participation was thought to arise from the extra
incentives (an iPod drawing in addition to extra credit) and email reminders used in study 2. As was previously mentioned, the experimental design for Chem 431 brought several validity concerns, testing interference (test-retest very close together in time) that favored the control group and a threat of maturation that favored the experimental group. Even so, the general trends for Chem 431 in study 1 (case 2) were found to the same as those from study 2 for all statistical tests. Thus, we believe that by applying these new procedures, significant results for the same variables (as study 2) would be obtained for Chem 431 as well, and so data from this course can be used to support the conclusions from this work. Reinforcing this, the independent t-test for confidence for class skills was found to be significant (E > C) for Chem 431 data. Coincidentally, the only measure found to be significant for this data set (study 1, case 2) was also the only measure that differed from those used for Chem 457 (because each course had different class skills).

4.3.3 Conclusions Relating to the Hypothesis and Research Questions

Conclusions relating to the hypothesis and research questions posed as part of this assessment are shown below (Figure 4.15). The proposed hypothesis was disproved. Depth of understanding did not lead to a detectable difference in problem solving scores for the HETCOR problem, and both the experimental and control groups had similar scores for this measure. Asieba\(^1\) has found that explicitly teaching both content and problem solving methods were necessary to see an effect on problem solving skills. Therefore adding a problem solving strategy component when teaching the NMR modules may be required in order to see an effect on this
variable.

While the proposed hypothesis was disproved, there is evidence to suggest that several of the premises it was based on are true. The NMR modules designed as part of the NSF grant did promote deeper understanding of NMR concepts. The modules also increased student confidence for not only the class skills practiced, but also for NMR tasks in general. The latter extension represents an added bonus for students completing the modules. Despite the fact that problem solving was not directly affected by this NMR curriculum, enabling students to think more deeply about chemistry concepts and increasing confidence for related tasks are sufficient reasons to continue using the NMR modules at Penn State University.
Figure 4.15. Summary of Findings for NMR Assessment Project. In the following, results of this assessment are related to the hypothesis and the research questions posed in Chapter 2 (section 2.2.5).

Implications of NMR Assessment Toward the
Hypothesis and Research Questions

1. The primary hypothesis that “deeper understanding of one particular analytical technique (NMR) will increase undergraduate students’ abilities to solve chemical problems” was not supported. If there is an effect, it is too small to be measured with the methods used in this assessment.

   Evidence:
   a. NMR modules that were shown to increase depth of understanding (see #2) did not affect problem solving. Problem solving was not significantly greater for experimental group for study 2. However, expected trends (E > C) were found for study 2 and both cases in study 1.
   b. Depth of understanding may still be a small contributor to problem solving skills. A moderate positive correlation was found between depth of understanding and problem solving for study 2.

2. NMR modules worked as intended and promoted depth of understanding of NMR concepts.

   Evidence:
   a. Depth of understanding was significantly greater for experimental group for study 2. Expected trends (E > C) were found for both cases in study 1.
   b. Depth of understanding significantly increased over the course of the experiment in study 2. Also, expected trends (Post > Pre) were found for both cases in study 1.
3. Confidence for NMR tasks was linked to both depth of understanding and problem solving skills. Therefore the suggestion that it may be an intermediary between the two seems plausible.

Evidence:

a. Moderate positive correlations between depth of understanding and both confidence for class skills and confidence for NMR tasks were found for study 2 data.

b. Weak-moderate positive correlations among problem solving and both confidence for class skills and confidence for NMR tasks were found for study 2.

4. Some types of task confidence are affected by the NMR modules, but only those related to NMR.

a. Both confidence for class skills and confidence for NMR tasks (not practiced in the course) were positively affected by completion of the NMR modules. Also, these two types of confidence were strongly correlated.

Evidence:

i. For Chem 457, confidence for class skills and confidence for NMR tasks mean scores were significantly greater for the experimental group in study 2 (E > C).

ii. For Chem 431 (study 1, case 1), confidence for class skills was significantly greater for the experimental group. The expected trend (E > C) was found for confidence for NMR tasks even though this result was not significant.

iii. Confidence for class skills and confidence for NMR tasks were strongly correlated for ALL data sets (studies 1 and 2).
b. Confidence for general science tasks was not affected by the NMR modules. This type of confidence was linked to other confidence measures, but the strength of these correlations was inconsistent among data sets.

Evidence:

i. Confidence for general science tasks scores were not significantly greater for the experimental group in either study. Trends were inconsistent among data sets. Study 1 case 1 (Chem 457 fall 2006) showed a trend (C > E) opposite from the other two data sets, though no results were significant.

ii. The dependent t-tests showed that the pretest and posttest scores were not significantly different for confidence for general science tasks, and trends were inconsistent among data sets. However, both study 2 and study 1 (case 2) showed the expected trends (Post > Pre).

When reviewing the results of this assessment, we expect that the positive outcomes that result from of this NMR curriculum will result in its continued use at Penn State. However, the strengths and weakness of this assessment are described in the next chapter. Where this assessment was found to be lacking, an alternate assessment plan is proposed to provide additional support for decision-making regarding this program’s future.
Chapter 5

Program Evaluation: Assessment from Another Perspective

5.1 What is Program Evaluation?

The assessment described in this work was done from the perspective of basic research, but when considering future directions of this project, we believe that using a method complimentary to our initial approach would result in additional insights into the research problem. Program evaluation\textsuperscript{83,80} offers a complimentary approach to creating assessments that is often used for educational programs (school- or district-wide programs), but which remains virtually unknown to the chemical education community at large. Using a basic research approach, we began from a hypothesis, designed an experiment, collected and analyzed data, and used these results to form conclusions. However, as has already been mentioned, studies that take place within the confines of the classroom can never be as tightly controlled as an experiment run in a lab. Program evaluation uses a slightly different approach to this problem. Rather than trying to control the classroom environment, this type of assessment is designed to measure and evaluate an educational program “as is”, in the context that the program will be used in. In program evaluations, results may or may not be based on theories, and conclusions are used as decision points for stakeholders in the program. Should the program be continued? Can it be improved upon? Should the program be expanded? These are
questions that a researcher conducting a program evaluation is most interested in answering.

Program evaluation focuses on the needs met by an educational, health, or social program. Posavac and Carey\textsuperscript{83} define program evaluation as

“a collection of methods [and] skills . . necessary to determine whether a human service is needed and likely to be used, whether the service is sufficiently intensive to meet the unmet needs identified, whether the service is offered as planned, and whether the service actually does help people in need at a reasonable cost without unacceptable side effects.”

A needs assessment is often a vital component of a good program evaluation,\textsuperscript{83} and requires identification of stakeholders and their needs. After this information is collected, we can begin to design measures that show whether or not these needs are being met, and this information can help determine whether or not a program is likely to be used. Implementation refers to how the program is put into practice and is another key factor in any program evaluation.\textsuperscript{52} Sometimes, implementation of a program does not exactly match the program administrators’ original intent or expectations. If observations reveal discrepancies between program goals and daily implementation, the program administrators (grant-writers) may want to address these issues before deciding whether or not the program “works”. When needs are not congruent among stakeholders, even a program working as designed can lead to unwanted side effects. By including the needs of more than just higher-level stakeholders in the study, it is more likely that these side effects will come to light.

Conclusions drawn from the data in a program evaluation may be formative
or summative. **Formative** evaluations\(^2\) are used to make improvements to the program. For example, an educational evaluation may reveal ways to increase efficiency or to better meet the needs of the teachers. A **summative** evaluation\(^2\) is used to decide whether or not the program will be ended. For example, at the end of a grant period, researchers can evaluate a new program to determine if the money was well used and whether it would be worthwhile to pursue funding from other sources. The type of evaluation (formative/summative) is determined *not* by the intent of those designing the assessment, but by the use of the data after they are collected.

Program evaluation is a popular method of creating assessments among educators, and has several advantages over assessment from a basic research perspective. While this method is more often used to assess large programs across school districts or even states, it can also be used to gain insights into changes within a particular school. Assessments designed from the perspective of research often change teaching strategies in response to a theory that researchers want to test. However, program evaluation is well-suited to testing educational practices already in place. In fact, it is recommended that a program evaluation not commence until a new curriculum has been in place for 1-2 years. Waiting to assess the program allows any problems with implementation to be addressed before spending the money on assessment. Also, the goals of program evaluation are in-line with the goals of funded chemical education projects. For example, when the NSF funds a new chemistry curriculum (or teaching module), they are looking for evidence of innovation and require assessments that help them judge whether or not their
money has been well-spent. For the Penn State NMR curriculum, data collected would be used to decide how well the project worked to improve student learning. The grant-writers could also use the data to decide whether or not the implementation at PSU could be used as the basis of a phase II CCLI grant, where funding is awarded to export the new curriculum to other universities.

5.2 Strengths and Weaknesses of the Previous Method of NMR Curriculum Assessment

5.2.1 Strengths of the Previous Assessment

The NMR curriculum assessment used at Penn State (2005-2007) was approached as an experiment, and the data were used to draw conclusions based on our research questions (see section 4.3). In many ways, this was a useful method of assessment. One strength of this study was the experimental design. Results obtained from comparing experimental and control groups within the same lab sections are more valid a time measurements (pretest vs. posttest of one sample) or comparisons made among samples at different schools or in different classes. The measures/instruments designed were shown to be reliable (Cronbach’s alpha), and could be useful in future assessments of NMR modules. Online testing was successful in collecting data and minimizing class time used for testing measures; however we found that that online testing had to be coupled with incentives (iPod drawing) and email reminders. Data analyses were completed using statistics, and these calculations ($t$-tests and correlations) provided convincing data that the NMR curriculum was affecting students in positive ways, even though our primary
hypothesis was not supported.

Results from the NMR curriculum assessment could also have implications outside of the courses studied, and so there is the potential to impact the chemical education community. In general, research-based assessments lend themselves to this kind of broader impact, and this may be an advantage over program evaluations, which are primarily used for internal decision making. specifically, the results of this research may further the study of problem solving in chemistry, learner characteristics within the context of chemical education studies, and NMR specific educational assessments. This research supports the idea that problem solving does not improve through content and opportunities alone. Also, task confidence measures were quite informative to this assessment and could offer chemical education researchers an alternative learner characteristic measure for their studies. Now that established measures have been provided, one could envision studies that explore this variable's connection to other important learner attributes (e.g. motivation). By increasing the visibility of alternative variables, perhaps the chemical education community can move away from attitude-only studies, which have dominated published “assessments” in the Journal of Chemical Education (arguably the most accessible journal used by chemical educators).

The majority of measures/instruments developed for this study are very specific to NMR concepts and instrumentation, however this may actually increase the likelihood of them being used outside of Penn State. The NSF offers funding to aid universities seeking new instrumentation, and NMR spectrometers are quite often the instrument being requested. From 1997-2003, approximately 67 proposals
were funded through the NSF's CCLI Adaptation and Implementation (A&I) track to purchase FT NMR spectrometers. Previously, proposals that simply stated a need for new instrumentation were considered, but there is a movement within the NSF to discontinue that type of funding. Instead the NSF intends to fund only those instrumentation acquisition grants that also offer new pedagogy and/or innovation, and these must be assessed. Assessment tools specific to NMR content are expected to be highly useful to those seeking funding through this route. Few measures specific to NMR exist in the current educational literature, perhaps because most educational researchers do not have sufficient access to the NMR expertise required to create these measures.

5.2.2 Weaknesses of the Previous Assessment

In general, there are several disadvantages to using a research-based assessment instead of a program evaluation. Research-based assessments rely heavily on quantitative methods to provide data, but this becomes a problem when the number of participants (n) is small. Statistical power increases as the sample size increases. While there are statistical tests suitable for small n (n < 50), smaller effect sizes may not be detected using these samples even if they are present. This was the subject of the discussion presented in section 4.3.2. In this assessment, the Chemical Spectroscopy course (Chem 426) was excluded from the assessment because the number of students completing the hands-on NMR module was so small. This module was offered as a special project, where students worked to build their own NMR probe. It was a unique contribution to the program that may
have offered valuable insights. However, statistical measures on such a small sample (approximately 4 students per semester) would have been meaningless.

Another disadvantage of research-based methods is that they are not well-suited for the “messiness” of many classroom environments (for example, courses where there is overlap of assignments and concepts). Chem 457 did not suffer from this problem, but Chem 431 is illustrative of this. For the latter course, overlapping assignments that required the use of the NMR spectrometer were found throughout most of the laboratory course, and a weaker experimental design was the result (see section 3.1.2). While statistical calculations can still be performed on data taken from courses where there is overlap of content and activities, making the argument that a new module is a primary contributor to any detected change is made more difficult. If the results of an assessment cannot be used to convince outside sources of its findings, then it would be a waste of money. The data obtained for Chem 431 (study 1, case 2) represented a compromise in that overlap was avoided, but at the cost of a stronger experimental design.

Using the methods of program evaluation for future studies of the NMR curricula at Penn State would address these weaknesses. Program evaluation is designed to measure educational processes within the context that they occur. By using qualitative methods such as interviews and focus groups for smaller courses, additional information could be obtained. Since program evaluation focuses on informing the stakeholders rather than convincing skeptics in the scientific community, it may be suitable to study Chem 431 as an on-going process throughout the entire semester. Through more direct questioning, information on what portions
of the course were attributed to different student outcomes could be obtained from the perspective of teaching faculty, TAs, and the students themselves. While this type of study may be less valid to some, it could still be used to inform chemical educators at other universities with similar courses. Using a program evaluation approach to assess the NMR curriculum at PSU could also enrich the information already obtained. The previous assessment was designed to measure only those variables deemed important by a few stakeholders, the researchers, instructors, and NMR experts. Indeed, all testing measures (instruments) were based on these contributor’s expectations. However, by examining the needs of all the relevant stakeholders other benefits of the curriculum may be discovered.

There are other areas of the previous assessment that could be improved, and a list of these can be found in Figure 5.1. The most important addition to assessment of the NMR curriculum at Penn State would be collection of data for Chem 426 (Chemical Spectroscopy). As one of the most innovative aspects of the NSF-funded NMR curriculum, distribution of a probe kit to other universities has been discussed. Evaluation of this NMR module coupled with access to kits of parts (for an NMR probe) might improve the likelihood of the probe-building module’s adoption at other schools with NMR spectrometers.
Figure 5.1. Suggested Areas of Improvement for Previous NMR Curriculum Assessment. In the figure below, weaknesses of the previous assessment are given, and these are followed by recommendations for improving the assessment of the new NMR curriculum at Penn State.

Recommendations to Improve the Assessment of the NMR Curriculum at Penn State

1. Chemical Spectroscopy course (Chem 426) was not evaluated.  
   Recommendation: Collect data through interviews.

2. For the depth of understanding (NMR) measure, used only one item for each aspect of depth was measured.  
   Recommendation: Add more items to measure each of the two aspects of this construct.

3. The HETCOR problem used for the problem solving measure had low student scores (both pretest and posttest).  
   Recommendation: Use another measure of problem solving in addition to (or to replace) the HETCOR problem. Another option would be to measure a different variable (see #5).

4. The hypothesis of this assessment was disproved, and there was no evidence of problem solving being affected by either depth of understanding or the NMR modules.  
   Recommendation: Add a teaching component that addresses problem solving methods or measure a different variable.

The instruments themselves could also be improved. The depth of understanding measure was created to measure two aspects of depth specific to
NMR, however only one item per aspect was used on the final measure. These were free response items meant to remain open for student responses and gave usable data. The reliability calculation most relevant to these items was interrater reliability, and this was found to be acceptable for both items and the item total (all Pearson’s $r > 0.8$). However, it may be possible to expand the scale to include more items for each aspect of depth, resulting in a more reliable (item analysis) measure overall. While the mean score on the HETCOR problem was moderate for Chem 431, Chem 457 students scores were low for both pretest and posttest scores. In most data sets, there were also extreme high scores. For any test that yields extreme scores, there is a possibility of regression towards the mean. This means that the few students that scored extremely high on the first test will most likely have scores closer to the mean on the next test. The same will be true for students with extremely low scores. Even if students’ “true” scores don’t change, regression towards the mean is likely to occur due to random fluctuations in student test-taking from event to event. For this reason, adding another measure related to problem solving or changing the variable measured could benefit future assessments.

The previous assessment showed no evidence that the hands-on NMR modules or concomitant depth of understanding of NMR had any effect on student’s problem solving skills. This leaves program administrator’s (grant-writers) two options. One would be to add another component to the modules that teaches students problem solving strategies as suggested by Asieba’s work. The other option is to measure a different variable altogether. Any new variable chosen should be
somehow related to depth of understanding and should also be important outcomes for students taking the course. A needs assessment, performed before testing begins, could be used to find examples of the latter outcomes.

5.3 Proposed Program Evaluation of the NMR Curriculum at PSU

Program

The program undergoing assessment using the program evaluation approach is the new NMR curriculum (funded by the NSF) that was integrated across the undergraduate chemistry curriculum at Penn State. Usually the first step of a program evaluation would be to provide a summary of the key points of the program as defined by administrators of the program. In this context, we will refer to the administrators of the program as the grant-writers and research team responsible for content and overseeing implementation of new NMR curriculum at Penn State. Since an initial assessment has already been performed, this step has already been achieved, and the researcher responsible for assessment is familiar with the goals of the project (see chapter 1 for more details of the program). As a reminder, the administrators of the program stated that the primary objectives were for students graduating from the program to be able to:

1. Synthesize complex information into a more cohesive view of NMR, which would require students to make connections within and among the various aspects of NMR (theory, application, and instrumentation).

2. Move past the novice’s perspective of NMR, and view the technique more
as an expert (in NMR) would.

3. Translate gains (in knowledge, skills, confidence, etc.) from advanced experiences in one specific area (NMR) to future experiences with other analytical techniques.

4. Be better prepared for their future careers, as professionals in chemistry-related fields.

While these goals were enlightening, there may still be a need to further define the program for future assessments, especially the program’s link to student outcomes. Since the previous assessment did not find evidence of a link between depth of understanding and problem solving, it is up to the administrators of the program to decide what related factors they expect their curriculum to have an effect on.

_Evaluation Purpose_

The purpose of this assessment is to collect data on the program that can be used by the client (program administrators) to make decisions regarding the program. The proposal that follows aims to provide a series of evaluative questions and information needs, and to suggest data collection methods that could yield a more comprehensive assessment of the program. This proposed program evaluation would be used to supplement information already gathered through the previous assessment (Chapters 2-4), and to improve upon the methods used there.

This program evaluation will follow the procedures and guidelines found in
the CIPP model. When evaluating any program it is important to consider four factors: Context, Input, Process, and Products, and this is where the CIPP model gets its name. The CIPP model of evaluation was first introduced by Stufflebeam 40 years ago, and it stressed the importance of considering multiple factors, not just final outcomes, for any evaluation. Since its inception, Stufflebeam and others have used this model for a variety of assessments, and its proven utility has made it a staple for those assessing programs. In the proposed program evaluation, a needs assessment will be included in the first phase of assessment, the context study.

Stakeholders

To provide a more comprehensive assessment, the needs of multiple stakeholders should be taken into account. The primary stakeholders for this program are: program administrators, the funding agency, teaching faculty, undergraduates participating in the program (students), and the chemistry and scientific communities. Program administrators were consulted regularly for the previous assessment, and their input will be considered equally important for the proposed program evaluation. The funding agency, the National Science Foundation (NSF), is another primary stakeholder. As contributor of all current funding, this government agency is looking to see that their money is used efficiently by the program to maximize benefits to all participants. In addition, this agency can decide to reward successful programs with additional funding to expand their program to other universities (in a phase II CCLI grant). Teaching faculty, specifically those chemistry professors teaching the courses in which the NMR materials and hands-

193
on modules are used, are more directly connected to implementation of the program than other stakeholders. Therefore, their input will be most valuable in this study. Undergraduate students participating in the program are meant to be the primary benefactors of this improvement in NMR teaching. It is their outcomes that are expected to be the most influential to the continued use (and possible expansion) of the program, and thus their needs must be considered. Support from the chemistry and scientific communities can alter how those outside the participating courses view the program, so their perspective will also be considered.

When grant-writers make “broader impact” statements, it is most often the chemistry and scientific communities that they expect to be impacting. One benefit of this program might be providing the chemistry/scientific community with better-trained employees and colleagues as students graduate from Penn State’s undergraduate chemistry curriculum with more hands-on experience with advanced instrumentation. For this assessment, the chemistry and scientific community will be defined as any teaching or research faculty outside the direct influence of the program. Information from the chemistry community could be collected on various levels, depending on opportunity and funding. Chemistry teaching faculty and research faculty at Penn State could be contacted directly by those assessing the program. In addition, an annual Conference on Undergraduate Research and Education in Nuclear Magnetic Resonance (NMR) has been hosted by Bucknell University for the past several years, and this would provide an excellent opportunity to gather information from teaching faculty interested in NMR outside of Penn State. Chemistry teaching and research faculty outside of Penn State could
also be given surveys at ACS meetings, where the program administrators are presenting talks and posters on this program. Similarly, science faculty could be contacted at various levels. Teaching and research faculty within the Eberly College of Science (PSU) could be contacted easily. Scientific conferences could be used as opportunities to administer surveys to teaching and research faculty outside the university.

**Responsibilities**

Regarding the program evaluation proposal, both the researcher designing and implementing the assessment (assessment leader) and client (program administrators) will be expected to be responsible for aspects of the assessment.

Responsibilities of the assessment leader will include:

1. Identifying appropriate constituent groups (stakeholders) from whom to collect data.
2. Identifying the questions to be answered through data collection.
3. Designing appropriate testing measures/instruments.
4. Designing appropriate data collection methods, so that data can be collected throughout the evaluation process.
5. Providing a method for compiling and reporting data to the client.
6. Maintaining clear lines of communication with the client/stakeholders.

Responsibilities of the client will include:

1. Supporting data collection through resource allocation.
2. Using the findings of this assessment to make decisions regarding the program and its implementation. Assessment findings may be used to
improve the program, and/or to report to funding agencies or other outside sources.

3. Maintaining clear lines of communication with the assessment leader including giving feedback on the impact of the assessment.

4. Assessment leader will require some data to be collected directly from program administrators, and therefore she will require availability of these persons for in-depth interviews (IDIs).

5.3.1 Context

The context of this assessment refers to the interactions between the program and the environment in which it is implemented. In this case, the new NMR curriculum has been implemented within the undergraduate chemistry program at Penn State University. The make-up of students and teaching faculty within the department could be expected to affect how this program is implemented. In addition, how the chemistry and scientific communities view the program could contribute to its impact outside of the university. Overall, context is important because the long-term success and effectiveness of the program may ultimately be determined by these environmental factors. While gathering this contextual information, we will also be assessing the needs of primary stakeholders. This information can be used to guide data collection, and also to suggest areas of improvement to the program administrators should the program be deemed unsuccessful in its current state. To supplement data already collected from program administrators (in the previous assessment), data will be collected from the following stakeholders to provide information on context: students, teaching faculty, chemistry and scientific communities, and the funding agency or agencies.
Students

Since positive student outcomes are the goal of the educational program, these contextual factors are expected to be very important to this assessment. Demographic information such as past academic achievement, socioeconomic factors, and racial and ethnic backgrounds are relevant because they may alter how this program would be received at a school of significantly different make-up. Information about student demographics will be collected through archival data and surveys. Focus groups will also be held to determine general trends of students’ attitudes and needs. The insights gathered through these focus groups will be used to guide assessment leader in creating a survey to measure these same factors, and this survey will be administered to all participating students. Specific data collection methods can be found in Appendix G.

Information Needs: The following questions are representative of those that will be addressed through the methods described above.

1. What is the baseline academic performance of students in participating courses?

2. What academic achievement trends might contribute to the effectiveness of this program?

3. What percentage of students identify with certain minority or ethnic groups?

4. What student demographic trends (e.g. socioeconomic status, major, etc.) might contribute either positively or negatively to the effectiveness of this program?

5. What student majors are represented in the courses studied? What might a student’s major contribute to the effectiveness of this program?
6. What career goals do participating students have?

7. How do students perceive the courses in which the program is being implemented?
   a. Is this course required or an elective for students? If it is an elective, how many other electives are offered as an alternative?
   b. How do they hope to benefit from the course?
   c. How might this course affect their future employment?

8. Are students currently (or planning to be) involved in student research outside of a course during their undergraduate career?
   a. What benefits would they hope to receive from doing research?
   b. How might student research affect their future employment?

9. How do students generally perceive the program (NMR modules and materials)?

*Teaching Faculty*

Teaching faculty (instructors) are key to implementation, and their cooperation is necessary for a successful program. These faculty are responsible for the daily functioning of the program, and so their input is extremely valuable. Teaching faculty demographics and teaching experience may directly impact how well students learn from these instructors. Teaching faculty may also have valuable insights into student learning in their course. Information about teaching faculty demographics will be collected through archival data and surveys. Because the number of faculty participating in this program is low (n = 3), in-depth interviews (IDI’s) are a quite feasible method of data collection. However, a focus group with all three faculty could be conducted instead. These will be used to provide detailed information regarding this aspect of the context. Specific data collection methods can be found in Appendix G.
Information Needs: The following questions are representative of those that will be addressed through the method described above.

1. Does the instructor identify with a certain minority or ethnic group?

2. What teaching faculty demographic trends might contribute either positively or negatively to the effectiveness of this program?

3. What is the instructor’s educational background?
   a. Schools at which he/she received degrees?
   b. Area of expertise (Ph. D. research)?

4. What is the instructor’s previous teaching experience (e.g. number of years of experience teaching at college-level)?

5. How do the teaching faculty view the needs of the students in their course?

6. What are the teaching faculty’s expectations of student outcomes?

7. How do the teaching faculty generally perceive the program?

8. How does the instructor plan to help implement the program?

Chemistry and Scientific Communities

The chemistry and scientific communities are expected to benefit from this program, and addressing their needs would provide evidence of the *broader impact* of this NMR curriculum. As previously mentioned, this community could be assessed on multiple levels, and are broadly defined as any chemistry or scientific faculty outside of the program being assessed. Groups chosen for data collection will depend on opportunities and funding available. Subject to funding, several focus groups composed of either chemistry and/or science faculty will be held to determine general trends of the community's attitudes and needs. The insights gathered through these focus groups will be used to guide assessment leader in creating a survey to measure
these same factors, and this survey will be administered to other members of the chemistry (or scientific) community at conferences as opportunity and funding allow. Specific data collection methods can be found in Appendix G.

*Information Needs:* The following questions could be asked of any or all the groups defined above as “chemistry and scientific communities.” These questions are representative of those that will be addressed through the methods described above.

1. How do these community members view the needs of undergraduate students in their field?
2. What are these community members’ expectations of student outcomes?
3. What do these community members want their future employees or colleagues to know when they graduate from an undergraduate program?
4. How do these community members view the role of advanced instrumentation in undergraduate education?
5. How do these community members generally perceive the program?
6. How do these community members feel a similar program could affect undergraduate education at their universities?
7. What do these community members see as possible limitations to implementation of a similar program at their university?

*Funding Agency*

The NSF is a major contributor to undergraduate curricula innovations in the United States, but they require evidence of how funding choices have lead to programs’ success. Therefore, it is important that a program provide evidence that money is being used effectively and with notable results. There is also another incentive to consider the needs of the NSF. Programs that are deemed successful can
apply for additional funding to expand their program to other universities (through a phase II CCLI grant). It is unlikely that data regarding these needs could be obtained through direct contact with program directors at the NSF. However, some data can be collected through archival data found on the agency’s website, specifically the website of the subdivision through which the grant was funded (CCLI). This information can be supplemented with data collected through IDIs with the NMR curriculum’s administrators at Penn State. This would include the program administrators’ knowledge of specific grants applied for and their perspective on what is expected of them. Specific data collection methods can be found in Appendix G.

*Information Needs:* The following questions are representative of those that will be addressed through the methods described above.

1. Under what branch of agency was this project funded?
2. Why was this program chosen to receive funding?
3. What outcomes are expected at the current funding level?
4. What type of assessment is required?
5. What outcomes and further information would be required for expansion of the program (phase II grant) to be considered?

*Interpretation of Results*

Contextual information will be used to provide valuable data on whether or not the program being assessed might be affected by the environment that it has been implemented in. If the program is found to be unsuccessful overall, data
already collected can be grouped in various ways to look for trends. For example, the assessment leader could determine whether chemistry majors scored significantly better on the outcomes measured than chemical engineering majors. Similarly, students from different racial or minority backgrounds, or different levels of academic achievement can be compared. Interactions between teaching faculty demographics and student demographics can also be identified.

5.3.2 Inputs

For this assessment, input refers to all the factors that go into a program’s implementation. This includes monetary contributions, resources available, and even training materials and procedures. Since all of these are necessary for a successful program, the assessment leader should be aware of all contributions. In addition, this information can be used in a formative manner. If a program is not working, the inputs can be examined to determine if more resources need to be allocated for specific parts of the programs.

Budgetary Contributions

Programs cost money, and the total budgetary contributions must be comparable to the minimum funds required to sustain the program. If this restriction cannot be met, then the program has no future. When funding is limited, it may be necessary to compromise aspects of the program. An evaluation of the budget and costs can be used to assess whether the use of funds is working as is, whether funds already obtained should be redistributed, or whether additional funds
are required. Information on budgetary contributions will be collected through IDIs with program administrators. Specific data collection methods can be found in Appendix G.

**Information Needs:** The following questions are representative of those that will be addressed through the methods described above.

1. What funds are available for at the current funding level:
   a. for developing NMR materials and hands-on NMR modules?
   b. for training teaching faculty?
   c. for assessment of the program?
2. What, if any, funding is available through sources outside of the funding agency (e.g. through the university)?
3. What funding is available for on-going implementation of the program once the grant period has ended?
4. What funding might be available if program administrators choose to expand the program?

**Resources Available and Costs of the Program**

The undergraduate chemistry program at Penn State offers many valuable resources. It is important to consider the resources needed not only for the NMR materials and modules themselves, but also for training of teaching faculty. The purpose of this portion of the assessment is to determine if these resources are being used efficiently, and whether ideal conditions for implementation are being met. Information on costs and resources expected for ideal implementation and training will be collected through IDIs with program administrators. Resources such as class time and instrument time will be discussed in IDIs with teaching faculty and instrument staff. Again this is made possible by the small number of contributing
faculty (three teaching faculty and one instrument staff person). Specific data collection methods can be found in Appendix G.

**Information Needs:** The following questions are representative of those that will be addressed through the methods described above.

1. What resources (training staff, time, etc.) are available for faculty training?

2. What are the expected monetary costs are associated with:
   a. developing NMR materials and hands-on NMR modules?
   b. training teaching faculty?
   c. assessment of the program?

3. What other resources are required for ideal:
   a. implementation of NMR materials and hands-on NMR modules? (e.g. class time, instrument time, computer time, instrumentation, instrument staff, etc.)
   b. training of teaching faculty? (e.g. training staff, time, incentives, etc.)

4. What resources are currently available for:
   a. implementation of NMR materials and hands-on NMR modules?
   b. training of teaching faculty?

5. What is expected of teaching faculty to support the implementation of this program?

**Teaching Faculty Support**

It is expected that teaching faculty will have to alter their courses to include new curricular materials and modules that focus on NMR. The level of their involvement and support of the program could be critical to its success or failure. If new curricular materials can be integrated into existing courses seamlessly, they are more likely to be utilized in the future. For these reasons, the assessment leader would want to know exactly what changes were being made to the previous course
procedures in order to sustain the program. Information on course procedures and daily implementation of the NMR curriculum will be collected through IDIs with teaching faculty and instrument staff. Specific data collection methods can be found in Appendix G.

*Information Needs:* The following questions are representative of those that will be addressed through the methods described above.

1. What was the previous course schedule?
2. How was the previous course schedule altered to include this program (NMR materials and hands-on NMR modules)?
3. How was the course syllabus altered for the program?
4. What class time is allotted for this program?
5. How does the allotted time differ from the *ideal* amount of class time suggested by the program administrators?
6. What is expected of you, as teaching faculty, to support the implementation of this program?

*Interpretation of Results*

Input information will be primarily used to determine if the current allocation of funds and resources are sufficient and whether they can be used to sustain the program in the future. This information can also be used to find weaknesses in program implementation, if the program is not working as desired. Input data can reveal if re-allocation of resources to certain aspects of the program is necessary to improve overall function.
5.3.3 Process

When an assessment team studies the “process” of a program, they are collecting information on implementation. The primary purpose of this aspect of the assessment is to determine if the program was implemented as designed. In addition, it can be used to see how program implementation might be affecting efficiency and outcomes. Any barriers to proper implementation should be revealed through this phase of the assessment.

Ideal Implementation and Program Administrator Support

In order to determine if the program has been implemented as designed, it will be necessary to consult the program administrators. The details of implementation may take time to work out, and so it is recommended that a program evaluation not commence until a new curriculum has been in place for 1-2 years. At this point, any problems with implementation should have already been addressed before spending the money to assess the program. However, it is still possible that the program has not been able to meet the ideal conditions, and the assessment leader should be aware of any deviations. Information of the ideal implementation will be collected through IDIs with the program administrators. This information will be used to formulate criteria for in-class behaviors that will be used for later data collection. In addition, information on the level of support offered by program administrators to teaching faculty will be assessed. Specific data collection methods can be found in Appendix G.
*Information Needs:* The following questions are representative of those that will be addressed through the methods described above.

1. **What teacher outcomes were expected to come from training?** For example:
   a. what level of program knowledge?
   b. what level of NMR content knowledge/expertise?
   c. what specific knowledge of implementation procedures?

2. **What specific behaviors would a well-trained teaching faculty for a course show during daily implementation of the NMR materials and hands-on modules?**

3. **What activities would the students be engaged in during proper daily implementation of the NMR materials and hands-on modules?**

4. **What specific behaviors would students exhibit during proper daily implementation of the NMR materials and hands-on modules?**

5. **What support offered by program administrators and/or the Mueller research team has been utilized by teaching faculty during implementation?** For example, how many times have the following occurred:
   a. answering questions related to proper implementation of the program?
   b. technical support for any problems encountered during hands-on NMR modules?
   c. answering questions on NMR content, offering expertise to teaching faculty?

*Teacher Training and Perceptions of Implementation*

The actual outcomes from training and for implementation of NMR materials is expected to vary from the expectations of program administrators. Therefore it will be necessary to find out what actually occurred from the perspective of those on the front-lines of implementation, the teaching faculty. Information on the teaching faculty’s knowledge of the program and perceptions of training will be gathered through surveys. These surveys will be designed using the program administrator’s
expectations a foundation. Other information to be collected through these surveys include instructors’ perceptions of implementation and access to support. Specific data collection methods can be found in Appendix G.

The teaching faculty training survey will be used to determine:

1. How program administrators expectations relate to actual:
   a. Teaching faculty’s level of program knowledge.
   b. Teaching faculty’s level of NMR content knowledge/expertise.
   c. Teaching faculty’s specific knowledge of implementation procedures.

2. Teaching faculty’s perceptions of their adherence to the program

3. Teaching faculty’s self efficacies in program implementation

4. How many times teaching faculty utilized support offered by program administrators and/or the Mueller research team during implementation. Specifically, how many times have they:
   a. asked questions/received answers related to proper implementation of the program?
   b. asked for/received technical support for any problems encountered during hands-on NMR modules?
   c. asked questions/received answers relating to NMR content beyond their current level of expertise of NMR?

**Implementation Fidelity Checks**

Rather than relying solely on stakeholders’ perceptions, the best way to definitively determine whether or not a program is being implemented as designed is to observe the program in action, meaning within the courses themselves. Without this confirmation, it may be possible that the ideal methods are never reaching their intended audience, the students. Information collected from program administrators will be used to design tools for observing both students and teaching faculty during
implementation. These observations can be performed by the assessment leader or
trained observers, depending on the funding available. All observation tools will be
made prior to actual observations. Specific data collection methods can be found in
Appendix G.

Classroom observations will be based on program administrator’s expectations.
Observers will be looking for evidence of:

1. Specific behaviors expected of a well-trained teaching faculty during daily
   implementation of the NMR materials and hands-on modules.

2. Activities expected for students engaged in proper daily implementation of
   the NMR materials and hands-on modules.

3. Specific behaviors expected of students during proper daily implementation
   of the NMR materials and hands-on modules.

Interpretation of Results

Investigation into the processes of the program will provide the assessment
leader with valuable information regarding the implementation of the program. If
implementation is close to the ideal, nothing more needs to be done. However, for
cases where direct observation does not reflect what was expected in an ideal
implementation, changes may need to be made to the program. Therefore, this
information can be used to make improvements to the program or attributions to
why it did not succeed.
5.3.4 Products

While it is important to consider the context, inputs, and processes of a program, the products (outcomes) of a program are undoubtedly given the most weight in any evaluation. The previous assessment focused only on products, and this section is where the two assessments will be the most similar. Where measures are the same, we will use suggestions provided in Figure 5.1 (section 5.2.2) to improve the instruments. By assessing the needs of other stakeholders, other outcomes to measure may be discovered. These would represent factors important not only to the program administrators, who were the only contributors to the outcomes measured in the previous assessment, but also other important stakeholders such as the chemistry community or the students themselves.

Student Outcomes

The most important difference in outcome measures between the two assessments will be problem solving. As was previously mentioned, since the previous assessment did not find evidence of a link between depth of understanding and problem solving, program administrators will have to decide what related factors they expect their curriculum to have an effect on. This may result in one of several options: further study of problem solving with the addition of problem solving components to NMR modules, further study of problem solving without changes to the curriculum, or study of a different variable. Adding an additional component to the NMR modules where problem solving is taught explicitly would be in-line with the recommendations found in Asieba’s chemical education study.
(which has been previously described). This would require changes to the curriculum as it is, and therefore these new components would have to be used within the curriculum for at least one course cycle to ensure that the methods of implementation were stable. An advantage of this option would be that the previously established problem solving task could be used. Testing and scoring methods for the HETCOR problem have already been established, so this would be a convenient measure. To confirm that all other effects of the modules remained the same, the other variables found to have an effect in the previous study (depth of understanding, confidence for NMR tasks, and confidence for class skills) should be administered again.

If program administrators still believe that their NMR curriculum should have a measureable effect on students' problem solving skills, and they wish to study this without changes to the curriculum, we suggest using another measure related to problem solving. The Science Process Skills Test (SPST) is designed to measure skills necessary for solving problems in science, specifically: identifying variables, identifying and stating hypotheses, operationally defining, designing investigations, and graphing and interpreting data. This test was originally developed by Burns, Okey, and Wise\textsuperscript{88} to measure skills across scientific disciplines, and it was designed to be used for middle and high school students. In this 36 item multiple choice test, students are given example problems/experiment set-ups and asked to actually perform one of the skills. For example, a study of auto efficiency is described, and then students are asked to state what information within the experiment was actually being used to measure auto efficiency (operational
definition). The reliability of this test, calculated across multiple samples ranging from 7th - 12th grade, was found to be 0.86 (Cronbach’s alpha). Validity measures also showed reasonable discrimination among items measuring different constructs. Even though this test was designed for middle and high school students, there is precedence for its use with undergraduates. In a study performed by Gurses et. al,\textsuperscript{89} student learning in a physical chemistry laboratory course was assessed using this measure. Results of this work showed that the Problem-Based Learning (PBL) based laboratory assignments being evaluated did promote an increase in students’ science process skills. Scores on the SPST were found to be significantly greater on the posttest when a paired-samples $t$-test was calculated ($t(39) = -4.66; p < 0.05$). This research was conducted in Turkey, but students tested were at a similar level (juniors) in their undergraduate chemistry curriculum to the population that we wish to study. Therefore, this measure might be useful for a study at Penn State. This could be used in lieu of an actual problem solving task or to supplement a new problem solving task. Since the modules themselves would not be changed in this scenario, there would be no need to re-test for variables used in the previous assessment.

The third option of measuring a variable other than problem solving would require no changes to the NMR curriculum, but new measures may need to be developed. Program administrators’ decision to measure another variable may be based either on their own expectations or other stakeholders needs. Regardless, measures would be developed by the assessment team in a similar manner to the problem solving measure. Relevant literature would be consulted, and any
adaptation required to make an instrument more NMR specific would be done through consultation with experts. In this case, there would be no need for measurement of the other variables used in the previous assessment (e.g. depth of understanding), since the NMR modules were not changed.

**Data Collection**

Information would be collected through online testing using the optimized methods and experimental designs for each course. Methods for Chem 457 (Physical Chemistry laboratory) and Chem 431 (Advanced Organic and Inorganic Synthesis) would remain the same as those used in previous assessment. For Chem 426 (Chemical Spectroscopy), the number of students participating in the program is very small, so a focus group would be used to collect data on perceived benefits offered through the NMR curriculum for this course, rather than a survey. Specific data collection methods can be found in Appendix G.

**Interpretation of Results**

Data collected in the product portion of the assessment are expected to factor heavily into decisions about the program. Statistically significant differences between the experimental and control groups would indicate a successful NMR module. Also, if variables within the assessment are related, this would be indicated by significant correlations. For Chem 426, quantitative data will not be collected, and so instead responses from focus groups that were found to be in-line with the desired student outcomes will be used to indicate a successful NMR module for this
Overall, a program evaluation’s worth lies in its ability to provide formative information. If a program is deemed unsuccessful as judged by the products (student outcomes), it is possible to find where problems in the program may have arisen. For example, data received from the process portion of the assessment may reveal a problem with implementation, or the data pertaining to inputs may show that additional resources are required. Also, using the CIPP model provides additional information that can yield positive results that would have not been considered by the program administrators. Thus, by considering multiple stakeholders needs, we can obtain a more comprehensive view of the assessed program.
Chapter 6

Conclusions

The purpose of this project was to assess an undergraduate NMR curriculum at Penn State University. Funding for the new curriculum came from the NSF’s Course, Curriculum, and Laboratory Implementation (CCLI) program, and this type of grant requires assessment. The goal of the NMR curriculum as stated in the grant was to provide NMR content in greater depth by integrating NMR modules throughout the curriculum in physical chemistry, instrumental, and organic chemistry laboratory courses. Allowing students to have hands-on contact with the NMR spectrometer and data was also a key portion of the new curriculum. As a chemical education project, the assessment was designed to follow the process of the scientific method. In this hypothesis-driven assessment, researchers wanted to know whether deeper understanding of one particular analytical technique (NMR) will increase undergraduate students’ abilities to solve chemical problems. Additional research questions that would be used to support this hypothesis and further assess the NMR curriculum were also studied.

Without a well-designed assessment plan, any data collected would be useless, thus considerable time was devoted to the design phase of the process. After key variables were defined, testing instruments were designed to measure these variables, using the educational literature to provide precedence when possible. The primary variables measured in this assessment were: depth of understanding of NMR, basic NMR knowledge, problem solving skills (HETCOR problem), confidence
for NMR tasks and general science tasks, and confidence for skills used in class (within the hands-on NMR modules). All testing for this assessment was conducted online, to minimize interference with course activities. Students were offered incentives for their participation, and all testing methods and materials were approved through PSU’s Institutional Review Board (IRB). Within each course, students were randomly assigned to an experimental or a control group. The experimental design (testing schedule) was chosen to offer maximum validity, but some modifications had to be made to avoid conflicts with other work in the laboratory courses tested (Chem 457 and Chem 431).

After conducting a pilot study that allowed for optimization of the instruments and testing procedures, data were collected in Penn State courses during the 2006-2007 school year. In study 1, small samples of students were tested in both Chem 457 (case 1) and Chem 431 (case 2). Researchers altered recruitment methods for testing of Chem 457 in the spring by offering an additional incentive (iPod drawing) and sending email reminders, and participation increased for study 2. After data were collected, scoring began. For free-response items, rubrics were constructed to aid in scoring, and each was found to have good interrater reliability (Pearson’s $r$ correlations > 0.8 for all). Also, reliability (item analysis) was tested for all measures that had multiple items, and Cronbach’s alpha was found to be above the cut-off ($\alpha > 0.8$) for most measures.

The assessment of the NMR curriculum at Penn State relied on quantitative methods and statistical calculations. Data were analyzed and descriptive statistics, independent $t$-tests between the experimental and control groups, and correlation
statistics were calculated for each variable. In addition, for those variables included on the pretest, dependent $t$-tests between pretest and posttest scores were also calculated. Descriptive statistics were used to explore general trends in the data. Independent $t$-tests were used to determine if differences in the experimental (E) and control (C) groups were significant. For some variables, the experimental group had significantly higher mean scores (E > C), giving evidence that the new NMR modules were positively impacting student scores for that group of students. Correlation calculations revealed possible links among variables, and the $r$ value of significant correlations indicated the strength of this link (low, moderate, or strong correlations). The dependent $t$-test was used to find significant differences in the mean pretest and posttest scores for several variables. When the posttest mean score is greater than that of the pretest (posttest > pretest), there is evidence that student scores are increasing over the testing period.

Before discussing the conclusions drawn from these data, it should be noted that results from the individual data sets yielded several important findings. Comparing results from Chem 457 in both studies, data from study 1 (case 1) was found to have noticeably lower mean scores and in some cases opposite trends for $t$-tests (E < C scores, pretest < posttest scores). Different results for the same course (Chem 457) in studies 1 and 2 suggests that participants for study 1 represent a self-selected sample. When comparing Chem 431 data (study 1, case 2) to that found in study 2 for Chem 457, the general trends for all variables were found to be the same. This suggests that by modifying the testing procedures used in study 2, we could increase participation and that similar results would be found for Chem 431. Thus,
data from this course can be used to support the conclusions drawn from this work.

The results of study 1 and study 2 were used to draw conclusions based on the hypothesis and research questions proposed in this work. The data collected through this assessment did not support the researchers’ hypothesis that increased depth of understanding (of NMR) would lead to better problem solving skills for students. No significant difference was found among the means of the experimental (E) and control (C) groups, even though the experimental mean was higher in all cases (E > C for all data sets). Despite this, there may be a link between these two variables. A moderate positive correlation was found between depth of understanding and problem solving for study 2 data. Thus it was concluded that if there is an effect, it is too small to be measured with the methods used in this assessment.

Even though the hypothesis was not supported, this assessment yielded several positive results regarding other research questions proposed in this work. The hands-on NMR modules were shown to be working as intended and increasing depth of understanding of NMR. We confirmed that depth of understanding was increasing for students who participated in the modules (E group) through significant results for the independent \( t \)-tests between the experimental and control groups (E > C) and dependent \( t \)-tests between pretest and posttest scores (posttest > pretest). It was also proposed that confidence for NMR tasks could be a learner characteristic that acted as an intermediary between depth of understanding and problem solving, and indeed these variables were found to be linked. Both confidence for class skills used within the hands-on modules and confidence for NMR tasks not
practiced within the course were found to have positive correlations between depth of understanding (moderate correlations) and problem solving (weak-moderate correlations) in study 2 data. These links are important because use of the hands-on NMR modules was found to positively affect confidence for NMR tasks (both confidence for class skills and confidence for NMR tasks). Study 2 showed that the experimental and control groups had significantly different means (E > C) for both of these variables. In Chem 431, students were tested for a different set of class skills (since they had a different hands-on NMR module), but calculation of the independent t-test for confidence for class skills confirmed that the mean score was significantly greater for the experimental group (E > C) for this course. Confidence for general science tasks was not affected by the NMR modules, and this may be because student confidence for these tasks is already stable for upper-level students. Evidence of this is provided by the dependent t-tests comparing pretest and posttest scores, which showed no significant difference between the means for either study. Also, all three data sets were found to have almost identical mean scores for this measure.

Results from the NMR curriculum assessment could also have implications outside of the courses studied, and so there is potential to impact the chemical education community. In addition to providing instruments/measures that could be used outside the university, the results of this research contribute to the study of problem solving in chemistry, learner characteristics within the context of chemical education studies, and NMR specific educational assessments (Figure 6.1). Valuable information was gathered through the current method of assessment for the NMR
curriculum. However, improvements could be made to the existing assessment, and an alternate assessment plan using methods common to program evaluations in the educational field that could supplement the information found in this study has been proposed (Chapter 5).
Areas of Broader Impact That the NMR Curriculum Assessment Might Effect

1. **This study provides an example of using literature-based educational research practices in a chemistry setting.** Most importantly, comparing an experimental and control group is a more valid method of designing an experiment (compared to other methods commonly used in today’s chemical education research). By following IRB protocols and discovering recruitment methods that increase participation, we ensured that our results were publishable.

2. **Instruments/measures developed for this study could be used in chemical education studies outside Penn State** (see # 4 and #5). Measures and scoring methods used within this assessment were found to be reliable.

3. This research supports the idea that **problem solving does not improve through content and opportunities alone.** Therefore, if chemistry departments want to improve the problem solving skills of their undergraduates, these should be explicitly taught.

4. **Task confidence measures offer chemical education researchers an alternative learner characteristic to study,** as they were found to be quite informative for this assessment. This might lead to a movement **away from** the attitude-only studies often found in chemical education research (which are especially prevalent among assessments published in the *Journal of Chemical Education*).

5. **Assessment tools specific to NMR content are expected to be highly useful to those seeking funding** to acquire NMR spectrometers through the NSF’s Adaptation and Implementation (A&I) track.
References Cited


(3) In a search of open positions advertised in Chemical and Engineering News over six month period (Sept 07 - Feb 08) found 4 positions for full-time chemical education faculty.


(7) A recent search (2003) of the Chemical Educator and J. of Chem. Educ. showed few examples of published NMR pedagogy for physical chemistry courses


(31) Upon completion of this study, some of the undergraduate chemistry courses changed course numbers. A map of these number changes can be accessed at the following website: http://www.chem.psu.edu/courses (accessed July 01, 2008).


(35) Revolution NMR (Fort Collins, CO) has been working with Dr. Karl Mueller to manufacture of a probe kit based on the module developed for Chem 426. For contact information, see: http://www.revolutionnmr.com/


(66) iPod Nano is a trademark of Apple, Inc.

(67) The rubric used for Basic NMR Knowledge items can be obtained through direct contact with either the author of this work, Dr. Kimberly Cossey (knc122@psu.edu), or her advisor on this project, Dr. Karl Mueller (ktm2@psu.edu).


(70) Statistical Package for the Social Sciences 16.0 (Mac); SPSS, Inc. 2007.


(86)    Stufflebeam, D. *Theories Into Practice* 1967, 6, 126-133.


APPENDIX A

Hands-on NMR Modules (for Chem 431 and Chem 457)

NMR Module for Chem 431W: XWIN Tutorial for NMR Data Output

Chem 431W NMR Software Training (XWIN)

XWIN is the software used for Bruker NMR spectrometers. The goal of this training is to take you through the process of working-up a H-NMR spectra, meaning making it look nice and getting the maximum amount of data from each spectrum you take. You will have practice phasing, integrating, picking peaks (to obtain coupling information), and printing spectra. Special info for non-H spectra is given on pg. 8.

BY THE END OF THIS TRAINING, YOU SHOULD HAVE:
1. A full spectrum (10ppm-0ppm) - should be the right size and phased, must have integrals and peak picking.
2. Close-ups of all peaks - should be the right size and phased, must have integrals and peak picking (everything except size will still be there from when you do the full spectrum).
3. Get these signed by Kim, Dr. Masters, or one of your TAs.

Opening XWIN. Log-on to one of the NMR processing computers (computer lab, left side. There are 3 of them.) If they are all full, but you see that someone is doing NON-NMR work on one, you may kick them off! Sign-on, then find the X-WIN NMR 3.1 icon and double-click to open the program.

First you might was to familiarize yourself with the layout of XWIN. Notice menus are at the top. All the buttons you need will be on the left side. In these instructions, buttons will be underlined. Also, at the bottom of the window is a line where you will type commands. In these instructions, “commands” will have quotes. Always hit enter after typing a command.

*Opening a file in XWIN*

1. On the top, far left of the screen is the File> menu. Click this, then Search.

2. For this tutorial, choose the following:
   \textbf{Directory} = d:/data \textbf{user} = chem 431 \textbf{name} = NMRtrain

   *this list is in alphabetical order, search here for your sample name

Do not do this during tutorial time!! But later,

**For your own samples**, when automation prints out the initial spectra, the \textbf{user} and \textbf{directory} information you need should be in the top left-hand corner of the print out. The \textbf{name} field is NOT the “sample name” you gave the automation, but is the \textbf{DATE} you took the spectrum (this is on the top right of your printout). Then choose \textbf{expt no} which is also on the top right of your printout. Also, on the top left of the printout will be your name, the experiment type (e.g proton) and deuterated solvent. Be careful not to take someone else’s spectra on accident.

\textbf{NOTE: Lost spectra!} If you forget to pick up your spectra the day you run it. Go in the computer lab (whitmore 206) and look immediately to your right as you enter. On the wall is a metal file organizer with course names. NMR spectra found for your course will be in the ‘Chem431’ slot. (If it is not there, you may
be able to find your expt no. by looking on the automation screen in the 400 room. Otherwise you must search all expt no. on the date you took your spectrum or Redo the whole run!

**Moving Around in XWIN**

For many tasks, you will need to change the view of the spectrum. Getting a close-up view, spreading peaks apart, increasing or decreasing the size of the peaks, etc.

To move around in your spectrum, use the following:

Expand to show full spectrum

Move to left

Move to right

Expand incrementally

Condense incrementally

INCREASE size of spectra X2

INCREASE size of spectra X8

DECREASE size of spectra X2

DECREASE size of spectra X8

**MOUSE/ select mode:**

Use the left button to go into a mode. The pointer will turn into a thicker down-arrow, we’ll call this *select mode*. Use middle button (scroll button) to select a peak or area. Left-click again to exit select mode.

**Try this:** In the regular mode (XWIN starts in this), left clicking in the spectral window will put you in select mode. Middle-click on the far left of your sample, then middle-click on the far right of your sample. This will change your viewing window.

*Phasing spectra*

making sure all peaks are above the baseline, and look symmetrical (see Fig 1)

Phasing is important, especially if you plan to integrate. Taking the integration of an unphased peak will give false values. Integration calculates the area under the peak to give you the number of nuclei (H’s in this case) that the peak corresponds to. Above the baseline this area is positive, but under the baseline the area is counted as negative. Thus the overall integration would be incorrect if all peaks are not in phase.

1. Click *phase* button. This will take you to the phase window, meaning you are now in a different mode and different buttons will appear on the left.
2. Click **biggest**. This chooses the biggest peak and uses it as a point to phase from. Make sure that you can see the bottom of the peaks well. Making the peaks bigger (using *2 button) may make it easier to see and phase.

3. Click the **ph0** and HOLD it. While holding move the mouse up and down until your spectra looks phased (see Fig. 1). All peaks should be above the baseline. Also, look at the left and right side of each peak, both should look the same (symmetrical). However, if you have more than one peak overlapping, these may be unsymmetrical and that is normal.

4. You may need to fine-tune the phasing by repeating step 3 with the **ph1** button.

   **Figure 1.** On the left is a portion of a spectrum before phasing (unphased). To the right is that SAME spectrum after phasing.

   ![Spectrum Before Phasing](image1)
   ![Spectrum After Phasing](image2)

5. When you are done phasing, click the **return** button. Then choose **save and return** button from the pop-up menu. Click OK.

   ***Calibrating ppm***

   Calibrate to your solvent peak. In most cases, CHCl$_3$.

   Here we will calibrate the shift ($\sigma$) to an internal standard, your deuterated solvent.

   **Question:** Why is it important to use a deuterated solvent for all liquid NMR samples?

   Answer: Two reasons. The most obvious is to think about what would happen if you use a non-deuterated sample, one with H’s on it. Because there is much more solvent than sample, the H’s of the solvent would give very large peaks which would cover and overwhelm your sample peaks. But there is another even more important reason . . .

   NMR spectrometers are tuned to the frequency of the nuclei they study, in this case H (for proton [H]) NMR, the frequency is very close to 400MHz. Carbon-13 would be very close to 100M Hz on the same spectrometer). In order to stay correctly calibrated, they relate this frequency to another value. Deuterium becomes this standard. Deuterium is NMR-active, and so can be used as a comparison. This is what ‘Lock’ does, and why it must be done for every experiment (automation does this for you). Without lock, your
spectra will come out unreadable.

This is why you should **ALWAYS USE DEUTERATED SOLVENT for NMR samples!! All deuterated solvents come from the STOCKROOM, except CDCl$_3$ (which is on the common shelf, in the brown bottle).** For Chem 431, it is common to make a product that is NOT soluble in CDCl$_3$. Think about what is in your sample and what solvent it is most like. This will help you choose a different NMR solvent (like dissolves like). Ask a TA for help if needed. Then go to the Stockroom for the **deuterated** solvent (do not use the solvents in the lab).

**Calibrating**

1. Look up value of your solvent on the chart, MSD Deuterated Solvents - Handy Reference Data. Shown below is a subset of this information (look at full list for any other solvents, found in computer lab). Note that some solvents have more than one peak.

<table>
<thead>
<tr>
<th>deuterated solvent</th>
<th>abbreviation</th>
<th>$\sigma$ (for H)</th>
<th>multiplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>chloroform-d</td>
<td>CDCl$_3$</td>
<td>7.26</td>
<td>1 = singlet</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma$ (for $^{13}$C) = 77.0</td>
<td>3 = triplet</td>
</tr>
<tr>
<td>acetone-d$_6$</td>
<td>d$_6$-acetone</td>
<td>2.04</td>
<td>5 = pentet</td>
</tr>
<tr>
<td>deuterium oxide</td>
<td>D$_2$O</td>
<td>4.65</td>
<td>br singlet</td>
</tr>
<tr>
<td>dimethylformamide-d$_7$</td>
<td>d$_7$-DMF</td>
<td>8.01, 2.91, 2.74</td>
<td>1, 5 = singlet</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 = pentet</td>
</tr>
<tr>
<td>dimethylsulfoxide-d$_6$</td>
<td>d$_6$-DMSO</td>
<td>2.49</td>
<td>5 = pentet</td>
</tr>
<tr>
<td>methyl alcohol-d$_4$</td>
<td></td>
<td>4.78, 3.30</td>
<td>1, 5 = singlet</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 = pentet</td>
</tr>
<tr>
<td>methylene chloride-d$_2$</td>
<td></td>
<td>5.32</td>
<td>3 = triplet</td>
</tr>
<tr>
<td>toluene-d$_4$</td>
<td>d$_4$-toluene</td>
<td>7.09, 7.00, 6.98</td>
<td>m = multiplet</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1, m = singlet</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.09</td>
<td>5 = pentet</td>
</tr>
</tbody>
</table>
Even though the peaks are given as the deuterated solvent, how would a deuterium peak show up on a proton (H) NMR? The answer is, it does not! What you are seeing is what happens when a D is exchanged with a H. For example, a solvent peak for d$_4$-methanol is at 3.30 ppm, but really this is for CD$_2$HOD.

**Label your solvent peak this way when annotating your spectrum.** Why is this peak a pentent? Because the H is coupling to the other 2 D’s on the carbon. Deuterium has a higher spin (\(I = 1\)) than hydrogen (\(I = \frac{1}{2}\)), so it splits more than a hydrogen would.

2. Find the solvent peak(s) on your spectrum. You may have to expand your spectrum to pick it out. Make sure the multiplicity (s, t, m, etc.) matches.

3. Click **Calibrate** button (on left side of screen)

4. Select peak and middle-click. Type in correct ppm from the chart. Hit enter.

*Integrating*

_tells you the ratio of hydrogens in each peak_

As previously mentioned, integration calculates the area under the peak to give you the ratio of nuclei (H’s in this case) that the peak corresponds to. It is important to remember this number is a ratio, not necessarily the actual number of H’s. If you calibrate your integrals to a peak (step 5) that you think is 1 H’s (and type this in) and it really is 2 H’s, other peaks on that molecule will come out to be half the true value.

**Question:** How can integration be used to obtain relative quantities of components of a mixture?

Answer: Let us consider an example scenario. Say you have a product from a reaction, call it X, and an impurity, call this Y. You are expecting to find X, therefore you know what peaks would arise from this molecule. Searching for a peak (from product X, if possible) that appears to be alone, with no other peaks overlapping, you find one. Knowing the structure, this peak should integrate to 1H. You calibrate this peak’s integral to 1.00. Now, all other peaks on product X should integrate to approximately whole numbers. They should do so in the ratio you expect from the structure of X. Other peaks in your spectrum, by process of elimination come from impurity Y (**REMEMBER: you may have more than one impurity, making elimination more complicated**). These have integrals of 0.33, 0.34, and 0.66 respectively. (What do you suspect about the ratio of X:Y at this point?) You now know that impurity Y is a molecule with 3 kinds of hydrogens, in a ratio of 1H:1H:2H. This combined with the shift information from these Y peaks, should help you identify it. Once you determine the structure of impurity Y, you are able to prove your suspicion that the ratio of X:Y is 3:1. **Make sure you understand how we figured this out!! Ask questions if you have trouble.** This process works the same if Y is a reactant, by-product, or solvent.

**Integrating**

1. Click **Integrate** button. This will take you into a different window.

2. Spread out the spectrum so that you can see the peaks well. (See Moving Around section for buttons.) Also you may need to increase the peak size. Make sure you can tell where one peak ends and another begins.

3. Go into **select mode**. Middle click on the left side, then the right side of a peak. Usually a number (the integral) will appear on the bottom, but if two peaks are really close together, it may not. This is fine, it will show up when you print.

4. Repeat step 3 for all peaks (except solvent peaks), moving around the spectrum as described above.

**NOTE:** Sometimes peaks of your sample overlap! This requires your judgement. If two peaks are barely
overlapping (the tailing side of each), it might be best to make it 2 integrals, cutting it off in the middle of the two peaks. This would result in a slightly lower integral value, but not cause a problem. However, if two peaks overlap to the point it is hard to distinguish where one ends and the other begins (e.g. Fig 1, phased spectrum, 5.8-6.0 ppm region), it would be much better to integrate the whole group. So if it would have been 1H and 3 H, now it will be one group as 4H. If you can still tell where the peak is centered (Fig 1, 5.83 ppm and 5.86 ppm), it may be acceptable to report these as separate peaks for your experimental section. Ask a TA if you are unsure.

**NOTE**: Do NOT integrate the deuterated-solvent peak, unless it overlaps with your sample peaks!

5. After integrating all peaks, look and see if they are all whole numbers (or very close to). For a pure compound they should be. If not, you should **calibrate your integrals.**
To do this:
   a. Find a peak that you know the number of H’s it should have. It’s ok to guess, if you have no idea what your compound is. For example, let’s say I have a triplet that I think should be 2 H’s.
   b. Put the cursor at the middle of that peak (my triplet) and left-click once. A white arrow should appear on top right of the integral curve.
   c. Click the calibrate button. Type the value you want. (I would type 2). Using the same ratios, all other peaks’ integral values will be calibrated to this one.

6. When you are done integrating, click return button (bottom left). Now click Save as ‘intrng’ and return on the pop-up menu. This will return you to regular mode.

**Peak Picking**

Peak picking shows peak labels you will need in order to determine coupling constants (J-values) for your peaks.

**Chem 431W**: YOU MUST DO THIS for Experimental Final Reports. All doublets, triplets, quartets on your H-NMR spectra must have reported J-values (in Hz) in the experimental section of your report.

1. Click the **utilities** button. This will take you to the utilities window.

2. Click the **MI** button. This is to set the minimum height you want it to label (tell you the Hz of). If an error pops up, just hit enter to obtain out of it.

3. A line should appear going horizontally across your spectral window. This line should be near the middle of the peaks that correspond to your compound. The idea is to keep it above noise & small impurity peaks, if you can. This keeps it from reporting every “blip” on the spectra, making reading the Hz values MUCH easier!

4. Left-click at the chosen level. For split peaks (triplet, quartets, and pentets especially) you need to make sure this line is below the outer end splitting.

5. Type “edg”. This will take you to the printing display window. Find PLABELS towards the bottom. Change this value to **yes** if necessary, by clicking the box. Then click ‘Save’.

6. To see if you have labels for all the peaks you want, preview the print-out by typing “view”, then hit enter. (Click ok if an error appears). This will show what peak labels will look like. Click Quit button to close window.
7. Leave utilities window by clicking the return button (bottom left).

**Printing**

1. Set spectral window to show full spectra, but less empty space.
   a. Click button that shows full spectra (see Moving around, above).
   b. Type “dpl1” (that is an L then a #1), and hit enter. Type in values you want – I usually use 10 (enter) then 0 (enter, enter). Make sure all peaks can be seen!! Some aldehydes and carboxylic acids have peaks >10ppm.

2. Print preview can be done by typing “view”, then hit enter. For full spectrum view, peaks should be as big as possible WITHOUT making the tallest sample peak go off-screen. If everything looks right (peak size is good, you see the integrals, peak labels, etc.), skip to step #4. If not, go to step #3.

3. To change the height of peaks, type “cy” then enter. A box will open up, change the value in the box to a higher number. It is fastest to change by multiples (e.g. if you want to double the size of the spectrum, double the value, change 12.5 to 25), then hit enter until the box closes. Type “view” then enter. If the peaks are still not big enough, repeat this step until they are.

   NOTE: If integrals or peak picking are not showing up: check and/or change printing parameters. Type “edg”, then hit enter. Look for INTEGRAL make sure it is set to yes. Look for PLABELS, make sure it is set to yes. Click SAVE.

4. To print, type “plot” and then hit enter. You are almost done, now go to step #5!

5. Print close-ups of your peaks! YOU MUST DO THIS FOR THE TUTORIAL.

   Also, This will save you time on assignments, so you can easily identify your peaks, and not have to come back later. Basically you want to make sure you can see the multiplicity of each peak (s, d, t, etc.) and see which peak label goes with which peak.
   a. Close-up to a group of peaks (usually 1-3ppm is a good size for a group) - In the regular mode (XWIN starts in this), left clicking in the spectral window will put you in select mode. Middle-click to the left of the group of peaks you want, then middle-click to the right of your group of peaks. This will change your viewing window.
   b. Type “dpl1” (that is an L then a #1), then hit enter. Hit enter 3X. This is setting the print window to be the same as the window you were just viewing.
   c. View by typing “view”, enter. All peaks should be at least 1/2 the size of the spectral area on the printout. If they are not, adjust “cy” (step #3) until they are.
   d. Type “plot” to print a close-up.
Special Requirements for $^{13}$C or other non-H spectra

all workup is the same EXCEPT

1. Check the MSD Deuterated Solvents - Handy Reference Data to find your deuterated solvent peak (found in computer lab). Notice the multiplicities may be different from what you would normally expect.

2. Do NOT integrate peaks. That is pointless for these kinds of spectra.

3. You DO need peak picking, but will use it to obtain a more precise shift (ppm) value. Type “edg”. Make sure PLABELS is set to yes. Now click on ETLABELS, change where it says Hz to ppm. Click Save. When you are done printing your spectrum and close-ups repeat this step and change it BACK to Hz, please!
Hands-on NMR Module for Chem 457: $T_1$ Experiment*

*Note: This procedure is a modified version of the experiment described in reference 1. Write-up provided by members of the Mueller group. Dr. Sergei Arzhantsev, instructor for the course, contributed questions for this assignment and a supplemental Hyperchem animation of hexanol.

Determining the Spin-lattice Relaxation ($T_1$) of 1-Hexanol using $^{13}$C-NMR

1. Objectives

- To determine the spin-lattice relaxation times ($T_1$) of each C atom of $n$-hexanol
- To relate $\tau_C$ values to the atomic motion of each C atom on hexanol

2. Introduction

Nuclear magnetic resonance (NMR) has become one of the primary tools in organic and biochemistry for structure elucidation, primarily through routine experiments detecting $^1$H, $^{13}$C, or $^{15}$N. However, NMR has the capability to far exceed the basics; almost every nuclei in the periodic table has at least one NMR active nuclei, and many possess spin 1/2, making them as straightforward to detect as $^{13}$C (figure 1).

![Figure 1. Nuclear magnetic resonance periodic table. Black squares contain atoms that have at least 1 nmr active isotope having spin 1/2, while white boxes contain atoms that have at least 1 NMR active nuclei having spin greater than 1/2 (quadrupolar nuclei).]
Experiments to determine inter-nuclear distance, through space coupling, coordination number and octahedral distortion are available through the use of NMR. This laboratory will focus on spin relaxation of a simple organic molecule measured through NMR. The data generated from this T1 experiment will be fit to a relaxation equation and converted to $\tau_C$, a variable related to atomic motion.

**Theoretical Background**

Nuclear magnetic resonance (NMR) spectroscopy exploits small energy differences in nuclear spin levels when spins are subjected to an external magnetic field (figure 2). This energy difference is described by the Boltzmann distribution,

$$\frac{N_{i,\text{upper}}}{N_{i,\text{lower}}} = e^{\left(-\frac{\gamma_i B_i h}{kT}\right)} \tag{1}$$

where $N$ represents the spin population of nuclei $i$ in either the upper or lower energy states, $\gamma_i$ is the gyromagnetic ratio of nuclei $i$, $B$ is the external magnetic field in units of Tesla (T), $h$ is Planck’s constant, $k$ is the Boltzmann constant and $T$ is the temperature. The population difference in a sample of H$_2$O in an external magnetic field of 9.4 Tesla is:

$$\gamma_{H} = \frac{-2.67519 \times 10^8 \text{rad}}{\text{sec} \cdot \text{T}}$$

$$\frac{N_{H,\text{upper}}}{N_{H,\text{lower}}} = e^{\left(-2.67519 \times 10^8 \text{rad} \cdot \text{s}^{-1} \cdot \text{T}^{-1} \cdot 9.4 \text{T} \cdot 6.626 \times 10^{-34} \text{J} \cdot \text{s} / 1.38 \times 10^{-23} \text{J} \cdot \text{K}^{-1} \cdot 293 \text{K}\right)} = 0.9999382$$

Since the signal intensity of all spectroscopic techniques relies on population differences this extremely small population difference has important consequences for NMR: the amount of sample required for NMR experiments is large, thereby increasing the absolute number of spins and increasing the absolute population difference. The presence of an external magnetic field causes the nuclear spins to align themselves with or against the applied magnetic field, dependant upon the initial energy state of the spin (figure 3) with
the energy difference between the two spin states given by $\Delta E = \gamma B_o h$.

Figure 2. Energy level diagram for a spin system.
The magnetization vector along the $z$-axis is given by the vector addition of magnetization of the $m = 1/2$ and the $m = -1/2$ (figure 3).

Figure 3. Magnetization vectors along $+z$ and $-z$ axis representing the two population states, followed by the resultant magnetization vector along the $+z$ axis generated by vector addition. However, in order to measure the difference and generate a spectrum the spin system is perturbed from equilibrium (magnetization along the $z$-axis) by application of an rf pulse (figure 4). The applied pulse has a tip-angle $\theta$, which is determined by the time length of the applied field is turned on (pulse length).
Figure 4. Application of a 90° pulse along the y-axis moves the magnetization vector from the z-axis to the x-axis. The tip angle is given by \( \theta \), the angle between the initial magnetization vector (\( I_z \)) along the z-axis and the magnetization vector after the applied pulse (\( I_x \)).

Once the applied field is removed the magnetization vector relaxes back to equilibrium along the z-axis. This relaxation effect is termed spin-lattice relaxation, or T1 relaxation (Figure 5).

Figure 5. Spin-lattice relaxation (T1 relaxation). Left: initial magnetization vector after a 90° pulse along y-axis; after a time period \( \tau_1 \), the magnetization vector along x has diminished and the magnetization along z is growing in. Right: equilibrium magnetization after time period \( \tau_2 \).

Spin-lattice relaxation occurs due to field fluctuations at the nucleus, and may be caused by:

- Magnetic dipole-dipole interactions
- Electric quadrupole interaction
- Spin-rotation interaction
- Scalar-coupling interaction
- Chemical shift anisotropy
Mathematically, T1 relaxation is described by the Bloch equation 2,

\[ \frac{\partial M_z}{\partial t} = -\gamma (M_x B_1 \sin(\omega t) + M_y B_1 \cos(\omega t)) - \left( \frac{M_z - M_0}{T_1} \right) \]  

The T1 decay may be modeled quantitatively by setting \( M_x = M_y = 0 \) in equation 2,

\[ \frac{dM_z}{dt} = -\left( \frac{M_z - M_0}{T_1} \right) \]  

(3)

Integration of equation 3 yields,

\[ M_z = M_0 (1 - 2e^{-\frac{t}{T_1}}) \]  

(4)

which will be used to model the T1 relaxation data.

### 6.1 Procedure/Data Analysis

Two pulse sequences may be used to acquire T1 relaxation data: the inversion recovery method, t1ir, or the T1 saturation recovery, t1sat. Note the two methods differ in their mathematical fitting equations, we will be using the T1 inversion recovery method (Figure 6).

**Figure 6.** Pulse sequence and fitting equation for T1 inversion recovery experiment.
How does T1 relate to Atomic Motion?

The T1 tells us the spin-lattice relaxation of a nuclei, in this case, different carbon-13 atoms, but it can be extended to give us useful information about the molecule. NMR active nuclei interact in many ways, but for this experiment it is the dipole-dipole coupling which is the most important interaction. This depends on the orientation and the distance between the two spins. (In this case, both $^{13}$C and $^1$H are NMR active and so they interact).

The dipolar interaction ($d$) is described by

$$d = \frac{\mu_0 h^2 \gamma_1 \gamma_2}{4 \pi r^3}$$  \hspace{1cm} (5)

where $\gamma_1$ & $\gamma_2$ = gyromagnetic ratio of the nuclei and $r$ = distance between atoms.

The T1 relaxation is proportional to $d^2/h^2$. As the carbon containing group (CH$_2$ or CH$_3$) rotates around its bonds, these distances between atoms on the group (nearby $^1$H’s) change, and these changes are transmitted to the carbon through coupling interactions. Thus it has been shown that T1 relates to motion of each group in the alkyl chain through a variable $\tau_C$, the effective correlation time for rotational reorientation, through the following equation.

$$\frac{1}{T_1} = N \left(\frac{\mu_0}{4\pi}\right)^2 \frac{h^2 \gamma_C^2 \gamma_H^2}{r_{CH}^6} \tau_C$$  \hspace{1cm} (6)

$$\gamma_C = 6.72881 \times 10^7 \text{ rad/(s} \cdot \text{T)} \hspace{1cm} \gamma_H = 2.67519 \times 10^8 \text{ rad/(s} \cdot \text{T)}$$

where $\gamma_C$ & $\gamma_H$ are the gyromagnetic ratio of $^{13}$C and $^1$H respectively, $\mu_0$ = permittivity in a vacuum, N = number of directly bonded H atoms (to the C of interest), and $r_{CH}$ is the C-H distance (in our case, the bond length for sp$^3$ carbon). Thus for liquids with low viscosity, such as hexanol, $\tau_C$ describes the average time for a the CH$_x$ group to rotate one unit angle.

The Experiment$^1$

For this experiment, you will be determining the T1 relaxation of all six
carbons of 1-hexanol. A degassed sample of 1-hexanol (70% in CDCl$_3$, sealed under vacuum) was placed in the spectrometer. This sample was tuned, meaning the NMR's two channels were tuned to the exact frequency of $^{13}$C and $^1$H respectively. The sample was locked to the frequency of the deuterated solvent (CDCl$_3$).

NMR spectrometers are tuned to the frequency of the nuclei they study, in this case $^{13}$C (the frequency is approximately 100 MHz on this spectrometer). In order to stay correctly calibrated, we relate this frequency to another value. Deuterium becomes this standard. Deuterium is NMR-active, and so can be used as a comparison. This is what 'Lock' does, and why it must be done for every experiment. Without lock, your spectra will come out unreadable.

Next the exact 90-times for this sample had to be determined for both $^{13}$C and $^1$H. This is the length of time that the spectrometer must pulse to rotate the magnetization vector 90 degrees. Hydrogen must be considered because the spectrum will be decoupled, and if the correct 90 time is not used for the $^1$H-channel, you could see splitting (coupling) of your $^{13}$C signal. You will be given these values in the procedure.

You will set-up and run the inversion-recovery experiment yourself on the 400 MHz Bruker spectrometer. The actual run time is approximately 35 mins. After this, you will process and analyze your data. Both running and processing your experiment will utilize XWIN NMR 3.1 software.

A T1 experiment is a pseudo-2D experiment, and it will give a 2D spectrum. In one direction (x-axis) you will be taking frequency information, this is the $^{13}$C-NMR spectrum for your sample. You must vary the delay time, $\tau$, this will become the y-axis of a 2D spectrum. (see below)
Figure 7. Pseudo-2D spectrum you will get from this experiment.

Fitting the plot of integral vs. t will give you a plot similar to that shown in Figure 6. This is an exponential function. You will be asked to CHOOSE your values to place in a vodelist. What is the best way to sample an exponential function? Where do you want to take the most points on the curve (Figure 6)? Think about this as you choose your values. The better your picks for the vodelist, the better your T1 fit to the data will be!

3. Laboratory Procedure

Experimental Set-up

1. Welcome to XWIN, a software program designed to run Bruker NMR spectrometers. The program should be open. This is the program you will be using both to run and to process your NMR T1 experiment.

2. First you will want to familiarize yourself with the layout of XWIN. Notice menus are at the top. In this procedure, Menus> will be written with arrows for the initial menu and the pull-down choices. (e.g File>Search). All the buttons you need later on to process will be on the left side. In these instructions, buttons will be underlined. Also, at the bottom of the window is a line where you will type commands. In these instructions, "commands" will have quotes. Always hit enter after typing a command!

3. "re filename"

The filename depends on your section and group, and should be in the form

chem457_s#_g%  where # = your section number and % = your group number (Don’t forget underscores!) For example, if you are in section 2, group 1 the filename would be:

cem457_s2_g1, and you would type “re chem457_s2_g1”

WRITE THIS FILENAME DOWN HERE: ____________________________
4. Choose the type of experiment you wish to run. For this experiment, you are running an inversion-recovery experiment. This has a pulse sequence of 180-τ-90.

“rpar” Choose: c13_Chem457_hex (include underscore). This file contains parameters you need to run the experiment.

5. Enter the 90-time for $^{13}$C by typing “p1”. Then enter the value 7.25. (This is 7.25 µsec)

6. You must also enter the 90-time for $^1$H since this experiment uses the proton channel for decoupling. Because it is the pulse length for decoupling, you must type “PCPD2”, then enter the value 109. (This is 109 µsec)

7. Between each pulse on the carbon channel, you must allow time for the spins to relax to zero (equilibrium). The common protocol is to wait AT LEAST 5 times the longest T1 in the sample. For hexanol, a delay of 30 secs is sufficient. Type “d1” then enter the value 30.

8. “edlist” Choose: vd on the pop-up menu. At the bottom of the pop-up window type your filename (see step 3).

9. This will open a notepad window. In this type 10 values. This is your vdlist (see intro), and it will contain your $\tau$ values. You will want to make sure you include more points at the curve of the function (Figure 6). To do this, values should be closer together then spread out as they increase.

Values of the vdlist must be typed in a certain format. Type one number per line. Do NOT include units - the program will assume the values are in SECONDS! Values less than 1 must have a zero in front of the decimal.

See example below for format, but DO NOT use these values, you MUST CHOOSE YOUR OWN using the instructions below!
Start with 0.20, hit enter, continue in the same manner with the next 8 points (your choice). This list is the \textit{vdlist} explained in the introduction (p. 6-7). Type points in \textit{increasing} order, so that the values start out close together but spread out, the further you go. For the final (10\textsuperscript{th}) point type 10.

\textbf{You must write down YOUR values in order (with the units) on this worksheet.}

10. Choose File>Save, then close the window (or File>exit).

11. type “vdlist” choose the name of the vd list you made in steps 8-10.

12. “eda” In the open window, type “sol” in the bar at the bottom of the window. Look for PROSOL and make sure the value directly below it says TRUE. If not, click it once, and it should change.

13. “ii” This double-checks your parameters for format errors. Wait until is says ‘ii finished’.

14. “rga” wait until is says ‘rga: finished’ at the bottom of the window.

15. “expt” This will tell you the length of your experiment. Make sure it is not too long!! It should be approximately 35 minutes. If it is much longer, see a TA for help.

16. “zg” Your experiment is now running. You can use this time to work on questions for the lab.

\textbf{Processing T1 Experiment}

1. Once your experiment is finished, you can begin processing. This experiment should be processed in the computer lab in Whitmore (rm 207). \{Data transfers from the NMR computer to the lab computers every 15 minutes.\}

2. \textbf{Opening XWIN.} Log-on to one of the NMR processing computers (computer lab, \textit{left} side. There are 3 of them.) If they are all full, but you see that someone is doing NON-NMR work on one, you may kick them off! Sign-on, then find the \texttt{XWIN NMR 3.1} icon and double-click to open the program.

3. On the top, far left of the screen is the File> menu. Click this, then \textbf{Search}.

4. For this experiment, choose the following:
   \begin{verbatim}
   Directory = d:/data       user = chem 457  *name = your filename
   *this list is in alphabetical order, search here for your sample name
   \end{verbatim}
   Double click the filename, and then click \texttt{apply}. Now click \texttt{close}

5. Look underneath the top menu and you should see your filename in the top left of the screen. The middle of the screen will be blank except for a message ‘type xfb to process’ Do NOT do this. Type “\texttt{xf2}”
6. "edt1" Change FCTTYPE to invrec, this tells the program what functions to use to calculate the T1 values.

Figure 8: The pseudo-2D spectrum. Note the full phasing key is shown.

7. Click the +/- button 2 times, to give the full rainbow of colors on the phasing key (see Figure 8). This should show both the positive and negatively phased peaks. If you cannot see any negative peaks, try clicking *2 to enlarge the spectrum on the top left several times (more peaks will appear). Similarly, /2 decreases the spectrum.

8. Click phase. The screen will become a split-screen. The top left window is your full 2D spectrum. All the buttons at the top left relate to this window. The 3 window on the right are for taking pieces of the spectrum. Buttons at the bottom left relate to these. (Figure 9).
Click in the full 2D window, now click the +/- button 2 times. You can also enlarge the spectrum similar to step 6.

9. Look at the peaks on the right side of the full spectrum (these are lower shifts, ppm). Click row, then middle-click on the lowest one (on the y-axis).

10. Move to position 1, by finding mov on the left side of the window, and clicking the 1 button. You should see a spectrum appear in box 1, the screen should look like Figure 9.

![Figure 9: Moving rows, in order to phase.](image)

11. Now on the original 2D spectrum, find the highest (on the y-axis) peak in the same column as you chose before. Click row, then middle-click the highest peak.

12. Move to position 2, by finding mov on the left side and clicking the 2 button.

13. Find big on the left. Click the 2 button next to it. (Figure 10)

14. Click the ph0 and HOLD it. While holding move the mouse up and down until your spectra looks phased (see Figure 11). All peaks should be above the baseline. Also, look at the left and right side of each peak, both should look the same (symmetrical). However, if you have more than one peak overlapping, these may be unsymmetrical and that is normal.
15. Use \texttt{ph1} (the same way as \texttt{ph0}) to make sure all peaks in each spectrum are phased.

\textbf{Figure 10:} Phasing your spectra
Figure 11. On the left is a portion of a spectrum before phasing (unphased). To the right is that SAME spectrum after phasing.

16. Click return, then save & return

17. Type “rspc”. This will give you a 1D spectrum of the first $\tau$ value from your vdlst. By integrating or picking peaks in this window, the software will automatically transfer this information to integrate or pick the same peaks for all 10 $\tau$ values from the vdlst!

18. Click integrate to enter a new window. Integrating will tell you the area under each peak. As $\tau$ decreases, this will decrease down to null, then become more negative. Thus we can use the changing integral values to monitor the change in magnetization as a function of the delay time.

19. Spread out the spectrum so that you can see the peaks well. (See Moving Around section below for buttons.) Also you may need to increase the peak size. Make sure you can tell where one peak ends and another begins.

**Moving Around in XWIN**

For many tasks, you will need to change the view of the spectrum. Getting a close-up view, spreading peaks apart, increasing or decreasing the size of the peaks, etc.

To move around in your spectrum, use the following:

Expand to show full spectrum  
Move to left  
Move to right  

Expand incrementally  
Condense incrementally
20. Click the left mouse button once. A white arrow should appear where the pointer is: this is called select mode. Middle click to the left side of (where it is still flat), then (where it becomes flat again) on the right side of a peak. Usually a number (the integral) will appear on the bottom.

21. Repeat step 19 for all 6 peaks (NOT the solvent peak, at 77.0 ppm), moving around the spectrum as described above.

22. When you are done integrating, click return button (bottom left). Now click Save as 'intrng' and return on the pop-up menu. This will return you to regular mode.

23. For the next step you will need to see all the peaks. First spread out the spectrum, so that you can clearly see the tip of each peak, by first clicking the expand button. Left clicking in the spectral window will put you in select mode. Now, middle-click to the left of your highest ppm peak in the sample, then middle-click to the right of your lowest ppm peak in the sample. This will change your viewing window.

24. Type "basl" to enter a new window. This is where you will tell the program which peaks it should find T1 values for.

25. Click def-pts on the left side of the screen (towards the top).

26. On the spectrum, left click once to enter select mode. For each of the 6 peaks (NOT the solvent peak!), middle click at the very tip of the peak. A small green arrow will appear to indicate a selected peaks. (see Figure 12).

27. When you have selected all 6 peaks, click return, and then save & return.

28. Click 2D in the bottom left corner.

29. “pd”

30. Analysis>relaxation (t1/t2). This will show the data for the first peak (highest shift, ppm).

Figure 12. basl Screen. Note the small arrow above a selected peak.
31. “ct1” to calculate the T1 value. It will show you the T1 curve for this peak. If you want, you can continue to view each curve by typing “nxtp” and repeating the “ct1” command.

32. You can get a printout of all the calculated T1 values, as well as the integrals for each value from your vlist, by typing “dat1”, and click ‘Print’. Turn in this printout with your final report.

4. **In-Lab Question**

1. The following is a plot of one doublet changing as a function of pulse length (t). Explain what is happening overall and at each point A, B and C. (All peaks ARE phased correctly.)

![Doublet plot](image)
5. **Data Analysis**

1. At the end of this report, a plot of the final $^{13}$C-NMR is given for 1-hexanol. **Label** each peak with the corresponding carbon (letters A-F). Ignore the solvent peak (a triplet at 77.0 ppm). Note: the shift of $C_C < $ shift of $C_D$, which is NOT what you would expect (this is a special exception for this gamma atom).

2. Using the data in your printouts, make a plot of $\tau$ versus intensity using a program other than XWIN. Apply a non-linear fit to these plots. Calculate the error.

3. Using the T1 data from XWIN, calculate the $\tau_C$ values for each peak. Make a chart with this information.

<table>
<thead>
<tr>
<th>Peak Letter</th>
<th>shift (ppm)</th>
<th>T1(s)</th>
<th>$\tau_C$ (ps)</th>
</tr>
</thead>
</table>

6. **Report Questions**

1. **Compare** the T1 data obtained XWIN to that of another program (data compiled in Data Analysis #2). What are the differences? What would account for these differences? Which do you think is more accurate and why?

2. Using the spectra and table made in data analysis (#1 & #3), **explain** the trend of $\tau_C$ values along the backbone of the 1-hexanol molecule. The molecule has been modeled and there are movies available in 207 Whitmore (see Dr. Arzhantsev for details). What do these tell you about the $\tau_C$ values?

References:

APPENDIX B

Final Posttest for Chem 457 and Chem 431

This survey was administered online according to the procedures described in Chapter 3 (section 3.3). Each section listed below represents one discreet webpage and the order was organized as shown.

Online Layout of Sections:

1. Basic NMR Knowledge & Depth of Understanding
   a. Part I: Basic Know., Reading Spectra Items (9)
   b. Part II: Basic Know., Definition Items (4)
   c. Part III: Depth of Understanding of NMR (2)
2. *Confidence for Class Skills: Chem 457 (9) or Chem 431 (10)
3. Problem Solving Task: HETCOR problem (4)
4. Confidence: MORE or LESS (7)
5. Confidence: General science and NMR Tasks
   a. Part I: Confidence for General Science Tasks (6)
   b. Part II: Confidence for NMR Tasks (5)
6. Attitude toward NMR (2)

* For the confidence for class skills section, students were shown ONLY those items relating to the course they were being tested in.

Section 1: Basic NMR Knowledge

Part I: Basic NMR Knowledge, Reading Spectra

For the following, we just want to know what you already understand about NMR. There is NO penalty for wrong answers.

1. For the following questions, use the $^1$H-NMR spectra (A-D) shown below.

![NMR Spectra]

a. Which spectra (A-D) is completely positively phased? (Hint: this is the spectra you usually see for common 1D experiments.)

b. Which spectra (A-D) is completely negatively phased?

2. For the following questions, use the $^1$H-NMR spectrum shown below.
Give the letter of the peak that: (letters may be repeated)

a. has the lowest shift (in ppm) ____

b. is a singlet ____

c. has the least hydrogens ____

d. is the farthest upfield ____

e. is coupled to another peak (there are TWO) ____ ____

3. For the spectrum shown above (question 2):
   a. How many unique hydrogens are there on the molecule? (How many different types of hydrogens?)

   b. How many total hydrogens are there on the molecule? (Give the minimum number possible).

**Part II: Basic NMR Knowledge, Definitions**

Remember: For the following: There is no penalty for wrong answers. We just want to know what you think!

In general, for a $^1$H-NMR spectrum:

4. What does a peak’s **integral** tell you about the corresponding hydrogens?

5. What does a peak’s **shift (ppm)** tell you about the corresponding hydrogens?
6. What does a peak’s **splitting pattern** tell you about the corresponding hydrogens?

7. What does a peak’s **coupling constant** \( J \), given in Hz \( \) tell you about the corresponding hydrogens?

**Part III: Depth of Understanding of NMR**

8. In general, what is NMR? What does it do?

9. There are many types of NMR experiments. What types of information about a molecule/reaction can you get using the NMR spectrometer? Please LIST as many as you can think of. (e.g. the structure of a molecule, etc.) Please do NOT give details about the experiment!
Section 2: Confidence for Class Skills

Class Skills Chem 457
This part of the questionnaire investigates the confidence you have in undertaking different tasks. Mark or circle the number that corresponds to your level of confidence. All are for NMR related tasks.

For example: If you do not feel very confident about taking an IR of your sample then you would answer 3.

Please indicate how confident you feel:

1. setting up an NMR experiment using X-WIN software
2. changing parameters for an NMR experiment (e.g. p1)
3. phasing for a T1 experiment
4. getting a T1 curve (τ vs. intensity) using X-WIN software
5. getting a T1 curve (τ vs. intensity) using another program (such as Excel or Mathematica)
6. using non-linear fitting on a T1 curve to obtain the T1 value using another program (such as Excel or Mathematica)
7. converting T1 to τc
8. relating τc values to atomic motion in a molecule
9. explaining what happens to the spins in a T1 experiment as τ is varied (what direction spins face, how the magnetization intensity changes, what happens to the peaks for each atom, etc.)

Class Skills Chem 431

Please indicate how confident you feel:

1. preparing a sample for NMR
2. setting up an experiment for NMR using the auto-sampler
3. setting up an experiment without using the auto-sampler (using X-WIN)
4. using a fourier transform (ft) to convert the raw data to a spectrum with peaks
5. phasing an NMR spectrum
6. integrating peaks (on an NMR spectrum)
7. printing a close-up of an area of peaks on your spectrum
8. determining coupling constants (J, in Hz) of peaks within your NMR spectrum
9. determining the identities of a mixture of compounds using NMR alone
10. determining ratios (relative amounts) of different compounds in a mixture using NMR alone
Section 3: Problem Solving Task, HETCOR Problem

NOTE for this section: PLEASE DO NOT GO BACK ONCE YOU HAVE ANSWERED A QUESTION! It is important for us to obtain your initial feedback as you proceed through the questions. There will be no penalty for incorrect answers.

Shown below is a HETCOR spectra. HETCOR is a 2D NMR experiment that relates peaks that arise from different types of atoms on same molecule. As you can see, the H-NMR spectrum is shown on the top and the C-NMR spectrum (for the same molecule) is shown on the left. The spots in the middle show which H-peaks go with which C-peaks. If a spot occurs, that means those atoms are attached directly. For example, one CH2 group would make one spot, with the shift of the 2 hydrogens in the x-direction and the shift of the carbon in the y-direction.

Use this HETCOR spectrum to answer the next 4 questions.

1. Using the HETCOR spectrum above, please list all the information you can figure out about the molecule. Assume the sample is pure (the solvent peak is labelled). Be sure to look at axes and units. Be as complete as possible. You do NOT need to explain WHY you believe what you do.
2. This question will vary slightly based on your course.
Try to list as many answers as you can think of. If you have not yet completed the experiment mentioned, just answer as best you can from the general knowledge that you have.

*Chem 431 students*: Please list all **similarities** between this HETCOR spectrum and the 1D H-NMR spectra you have processed on the 400-NMR.

*Chem 457 students*: Please list all **similarities** between this HETCOR spectrum and a T1 experiment.

3. This question will vary slightly based on your course.
Try to list as many answers as you can think of. If you have not yet completed the experiment mentioned, just answer as best you can from the general knowledge that you have.

*Chem 431 students*: Please list all **differences** between this HETCOR spectrum and the 1D H-NMR spectra you have processed on the 400-NMR.

*Chem 457 students*: Please list all **differences** between this HETCOR spectrum and a T1 experiment.

4. Using the HETCOR spectrum above, please list ANY ADDITIONAL information you may have figured out about the molecule by comparing it to NMR experiments you have already seen. (If no new information was gathered, put NONE).
Section 4: Confidence More or Less

MORE OR LESS CONFIDENT
Think back to the first day of this course. Do you think your confidence using NMR has improved in the following areas. Please rank on a scale of (Much less confident, less, slightly, same, slightly more, more, much more)

a. using the instrument itself
b. changing an experiment's parameters to obtain the best NMR data
c. processing raw data to obtain a spectra with peaks (1D or 2D)
d. printing out a nice looking spectra interpreting the structure of a sample from a spectra
e. determining the identity of a pure compound using NMR alone
f. annotating peaks on an NMR spectrum, if you already know the structure
g. getting the T1 from a peak on an NMR spectra

Section 5: Confidence for General Science and NMR Tasks
This part of the questionnaire investigates the confidence you have in undertaking different tasks. Mark or circle the number that corresponds to your level of confidence.

For example: If you do not feel very confident about talking to a scientist about chemistry then you would answer the following questions as shown:

a. Please indicate how confident you feel about talking to a scientist about chemistry
   Not confident  1 2 3 4 5 6 7 Totally confident

Part I: Confidence for General Science Tasks

Please indicate how confident you feel:

1. Ensuring that data obtained from an experiment is accurate Not confident  1 2 3 4 5 6 7 Totally confident
2. Proposing a meaningful question that could be answered experimentally Not confident  1 2 3 4 5 6 7 Totally confident
3. Explaining something that you learned in a chemistry course to another person Not confident  1 2 3 4 5 6 7 Totally confident
4. Knowing how to convert the data obtained in a chemistry experiment into a result that you could explain to a classmate Not confident  1 2 3 4 5 6 7 Totally confident
5. After reading an article about a chemistry experiment, writing a summary of the main points Not confident  1 2 3 4 5 6 7 Totally confident
6. Writing up the experimental procedures in Not confident  1 2 3 4 5 6 7 Totally confident
Part II: Confidence for NMR tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Confidence Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Explaining basic NMR theory to a friend not in this course</td>
<td>Not confident</td>
</tr>
<tr>
<td>8. Running an experiment on the NMR spectrometer without guidance</td>
<td></td>
</tr>
<tr>
<td>9. Designing an NMR experiment to answer a specific research question</td>
<td></td>
</tr>
<tr>
<td>10. Providing a correct scientific interpretation of NMR data</td>
<td></td>
</tr>
<tr>
<td>11. Optimizing conditions for an NMR experiment (e.g. tuning, shimming, setting 90 &amp; 180 times, setting sweep width, receiver gain, number of data points and scans. etc.)</td>
<td></td>
</tr>
</tbody>
</table>

Section 6: NMR attitude

10. How do you think the experience you gain by learning NMR techniques might affect the way to approach and/or understand a new analytical instrument?

11. You were previously asked about your goals for this course. Have you met these goals? Please list goals and explain why/why not. Also include any goals you may have added since the beginning of the semester.
APPENDIX C

Final Pretest for Chem 457 and Chem 431

This survey was administered online according to the procedures described in Chapter 3 (section 3.3). Each section listed below represents one discreet webpage and the order was organized as shown.

Online Layout of Sections:

<table>
<thead>
<tr>
<th>Grouping</th>
<th>(# items)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Scientific Interest and Previous NMR Experience</td>
</tr>
<tr>
<td></td>
<td>a. part I: Major and Career Goals (2)</td>
</tr>
<tr>
<td></td>
<td>b. part II: Previous NMR Experience (3)</td>
</tr>
<tr>
<td>8.</td>
<td>Basic NMR Knowledge &amp; Depth of Understanding</td>
</tr>
<tr>
<td></td>
<td>a. part I: Basic Know., Reading Spectra Items (9)</td>
</tr>
<tr>
<td></td>
<td>b. part II: Basic Know., Definition Items (4)</td>
</tr>
<tr>
<td></td>
<td>c. part III: Depth of Understanding of NMR (2)</td>
</tr>
<tr>
<td>9.</td>
<td>NMR Attitude (2)</td>
</tr>
<tr>
<td>10.</td>
<td>Confidence: General science and NMR Tasks</td>
</tr>
<tr>
<td></td>
<td>a. part I: Confidence for General Science Tasks (6)</td>
</tr>
<tr>
<td></td>
<td>b. part II: Confidence for NMR Tasks (5)</td>
</tr>
</tbody>
</table>

Section 1: Scientific Interest and Previous NMR Experience

Part I: Major and Career Goals

12. Which BEST describes your current major/concentration? Be sure to read all choices - some are similar.

a. chemistry
b. engineer: chemical or bio
c. engineer: mechanical, electrical, etc.
d. biology: biochemistry, molecular biology, etc.
e. biology: life sciences, premedicine, etc.
f. biology: animal science, agricultural science, environmental science, etc.
g. physical sciences & mathematics: physics, mathematics, geoscience, etc.
h. social/behavioral science: forensics, psychology, law, etc.
i. Other (please list):

13. Which best describes your particular area of scientific interest for your future career? (Base your answer on how you feel right now.)

a. basic research in academia
b. applied research in academia
c. basic research in industry
d. applied research in industry
e. professional degree in law
f. professional degree in medicine
g. management in industry
h. consultant
i. undecided
j. Other (please list):
Part II: Previous NMR Experience

14. Please list any courses you have taken that covered NMR topics prior to (or at the same time as) this course. List the course name and number where NMR was taught.

15. If you have taken courses which used NMR spectrometers, check each NMR instrument that was taught/used. Those shown below represent the NMR instruments in Whitmore, UP campus. Indicate N/A (not applicable) if did not get NMR training/practice.

   a. N/A
   
   b. 80 MHz - there are two of these found in the Instrument Room (Whitmore)
   
   c. 400 MHz - this instrument is in the small locked room and it has an auto-sampler
   
   d. Some other NMR spectrometer

16. Please list NMR training received (what techniques. e.g. H-NMR, C-NMR, DEPT, COSY, NOESY, etc.)
Section 2: Basic NMR Knowledge

Part I: Basic NMR Knowledge, Reading Spectra
For the following, we just want to know what you already understand about NMR. There is NO penalty for wrong answers.

1. For the following questions, use the $^1$H-NMR spectra (A-D) shown below.

![NMR spectra A-D](image)

a. Which spectra (A-D) is completely *positively* phased? (Hint: this is the spectra you usually see for common 1D experiments.)

b. Which spectra (A-D) is completely *negatively* phased?

2. For the following questions, use the $^1$H-NMR spectrum shown below.

![NMR spectrum B](image)

Give the letter of the peak that: (letters may be repeated)

a. has the lowest shift (in ppm) ___

b. is a singlet ___

c. has the least hydrogens ___

d. is the farthest upfield ___
e. is coupled to another peak (there are TWO) ___ ___ 

3. For the spectrum shown above (question 2):
   a. How many unique hydrogens are there on the molecule? (How many different types of hydrogens?)
   
   b. How many total hydrogens are there on the molecule? (Give the minimum number possible).

**Part II: Basic NMR Knowledge, Definitions**

Remember: For the following: There is no penalty for wrong answers. We just want to know what you think!

In general, for a \(^1\)H-NMR spectrum:

4. What does a peak’s **integral** tell you about the corresponding hydrogens?

5. What does a peak’s **shift (ppm)** tell you about the corresponding hydrogens?

6. What does a peak’s **splitting pattern** tell you about the corresponding hydrogens?

7. What does a peak’s **coupling constant (J, given in Hz)** tell you about the corresponding hydrogens?

**Part III: Depth of Understanding of NMR**

8. In general, what is NMR? What does it do?

9. There are many types of NMR experiments. What types of information about a molecule/reaction can you get using the NMR spectrometer? Please LIST as many as you can think of. (e.g. the structure of a molecule, etc.) Please do NOT give details about the experiment!
Section 3: NMR Attitude

1. How do you think the experience you gain by learning NMR techniques might affect the way to approach and/or understand a new analytical instrument (IR, UV-vis, GC/MS, etc.)?

2. What are your personal goals in taking this course?

Section 4: Confidence for General Science and NMR Tasks

This part of the questionnaire investigates the confidence you have in undertaking different tasks. Mark or circle the number that corresponds to your level of confidence.

For example: If you do not feel very confident about talking to a scientist about chemistry then you would answer the following questions as shown:

a. Please indicate how confident you feel about talking to a scientist about chemistry
   Not confident  1  2  3  4  5  6  7  Totally confident

Part I: Confidence for General Science Tasks

Please indicate how confident you feel:

1. Ensuring that data obtained from an experiment is accurate
   Not confident  1  2  3  4  5  6  7  Totally confident

2. Proposing a meaningful question that could be answered experimentally
   Not confident  1  2  3  4  5  6  7  Totally confident

3. Explaining something that you learned in a chemistry course to another person
   Not confident  1  2  3  4  5  6  7  Totally confident

4. Knowing how to convert the data obtained in a chemistry experiment into a result that you could explain to a classmate
   Not confident  1  2  3  4  5  6  7  Totally confident

5. After reading an article about a chemistry experiment, writing a summary of the main points
   Not confident  1  2  3  4  5  6  7  Totally confident

6. Writing up the experimental procedures in a laboratory report
   Not confident  1  2  3  4  5  6  7  Totally confident
Part II: Confidence for NMR tasks

7. Explaining basic NMR theory to a friend not in this course
   Not confident 1 2 3 4 5 6 7 Totally confident

8. Running an experiment on the NMR spectrometer without guidance
   Not confident 1 2 3 4 5 6 7 Totally confident

9. Designing an NMR experiment to answer a specific research question
   Not confident 1 2 3 4 5 6 7 Totally confident

10. Providing a correct scientific interpretation of NMR data
    Not confident 1 2 3 4 5 6 7 Totally confident

11. Optimizing conditions for an NMR experiment (e.g. tuning, shimming, setting 90 & 180 times, setting sweep width, receiver gain, number of data points and scans. etc.)
    Not confident 1 2 3 4 5 6 7 Totally confident
APPENDIX D

IRB Informed Consent and Recruitment Materials

Section 1. IRB 21504 Consent Form (Spring 07)

INFORMED CONSENT FORM FOR SOCIAL SCIENCE RESEARCH
The Pennsylvania State University

Title of Project: Evaluation of the Integration of Liquid and Solid State Nuclear Magnetic Resonance Spectroscopy (NMR) Into an Undergraduate Physical Chemistry Curriculum

Principal Investigator: Kimberly Cossey
### Chemistry Building, University Park, PA 16802
phone: (814) ###-####; knc122@psu.edu

Advisor: Dr. Karl T. Mueller, ktm2@psu.edu
### Chemistry Building; phone: ###-####

1. **Purpose of the Study:** The purpose of this study is to examine your perceptions of this course (Chem 457 OR Chem 431W). In particular, we want to understand what you like and don’t like about the course. Also, we are interested in your confidence level for specific tasks related to the course. The information collected in this project will be used to improve future sections of this course (Chem 457 OR Chem 431W). In exchange for your participation in this study, you will receive extra credit and may be entered in a random drawing for an IPOD nano.

2. **Procedures to be followed:**
   - If you agree to take part in this research, you will be asked to complete an online survey twice during the semester. The survey should take no more than 30-50 minutes each time. This survey will ask questions about how you feel the activities in the course affected your learning. The survey will be given online and you will be given a code number that will be used to track your survey. You have the right to not answer specific questions on the survey.
   - If you agree to take part in this research, you are giving the researchers permission to make copies of your course assignments and test responses and use them for research purposes. This information will be used to evaluate the activities in the course. You are giving permission for your course grade to be used as part of the study.

3. **Discomforts and Risks:** There are no risks in participating in this research beyond those experienced in everyday life. Sharing one’s assignments and grades may result in some discomfort. However, every precaution will be taken to ensure that only the researchers will have access to this information.

4. **Benefits:** The information collected will help to improve future sections of this course (Chem 457 OR Chem 431W). This research will add to the pool of knowledge about teaching chemistry courses.
5. **Duration/Time:** It will take no more than **30-50 minutes** to complete each online survey.

6. **Statement of Confidentiality:** None of the investigators involved in this study are instructors for this course. The investigators will be the only persons with access to this signed informed consent form. Your instructor will not know who participated in this study, and he/she will not know your responses to the surveys. Your grade in the course will NOT be affected by participation or non-participation in this study. Your student ID number on this consent form, answer sheets from course assignments, and exams will be used for tracking only. Your student ID number will be replaced with a code number. Only the researchers will have access to that information. All data will be kept in a locked filing cabinet for a one-year. The data will be destroyed after one year. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Dept. of Health and Human Services; The Penn State University Social Science Institutional Review Board (IRB); The Penn State University Office for Research Protections.

*Online Confidentiality:* We have implemented all available technologies to protect your confidentiality; however, Internet security has its limitations and no absolute guarantees can be made that a third party cannot gain access to the data.

7. **Right to Ask Questions:** You may ask questions about the research at any time by contacting:
   
a. Kimberly Cossey, Research Assistant in the Department of Chemistry; by calling (814) ###-####; or by emailing knc122@psu.edu

   b. If you have questions about your rights as a research participant, contact Penn State’s Office for Research Protections at (814) 865-1775.

8. **Compensation:** For each survey completed, you will receive **15 extra credit** points added to your final grade in this course. You will receive full points based on the completeness of your survey, NOT on the correctness of your answers. If you complete the survey, you will receive full points. In addition, those who complete **BOTH** online surveys will automatically be entered in a **random drawing for an IPOD nano.** One name will be drawn at the end of the semester and the winner will be notified by email.

9. **Voluntary Participation:** Participation is voluntary. Participants can withdraw from the study at any time by notifying the principal investigator. Participants can decline to answer specific questions.

You must be 18 years of age or older to consent to participate in this research study.

**This form will be accessible only to project investigators** will also be kept in a locked file in the Chemistry Building, Penn State University.

Please check the box beside the following statement to indicate if you wish to participate:

☐ I agree to participate in this study and give my informed consent.

---

**Signature:** ___________________________  **Date:** ________________

**Name (Please Print):** ___________________________
Email address: ________________

This Course Number (circle one):  

Chem 457  
Chem 431W
Section 2. IRB 21504 Script (Spring 07)

This semester there will be several opportunities to receive extra credit, as well as a chance to win an IPOD nano. By participating in a research study* about his course, you can help us to improve this course. In return, you will receive 15 points added to your overall course score for each activity completed. Due dates can be found on the handout given. Those who complete both activities will automatically be entered in a random drawing for an IPOD nano. Participation in this study is completely voluntary, and not participating will not affect your grade.

*To participate in the study and receive credit, you MUST sign an Informed Consent Form, which I will handout in a moment.

This research study is being conducted in affiliation with Penn State University. You must be at least 18 years old to participate.

Alternative: If you do not wish to participate in the surveys, you can alternatively receive an equal amount of extra credit by writing a 1 page review of an article from a scientific journal (email Kim Cossey, knel122@psu.edu for details). These article reviews would be due on the same dates as the corresponding extra credit activities, and be worth 15 points each.

NOTE: You cannot “double-up” extra credit by completing both an activity and a report in the same assigned period.

So for those of you interested, I need to tell you a little bit about the study.

We are conducting a research study about your perceptions of this course. In particular, we want to understand what you like and do NOT like about the course. Also, we are interested in your confidence level for specific tasks related to the course. By participating, you would be contributing to improvement of the course for future students, and even more so to the pool of knowledge about teaching physical chemistry. In addition to extra credit, for some parts of surveys, you will be practicing skills that may be useful in future assignments in this course!

If you chose to participate, you will be asked to complete TWO online surveys which should take no more than 30-50 minutes each. These can be found at the website found on the handout, I have given you. Surveys will ask questions on your perceptions of certain aspects of the course. In addition, you will be asked to answer questions related to certain topics within the course. It is important to note that you will receive extra credit points based on the completeness of surveys, NOT on the correctness of your answers. Meaning if you complete a survey, you will receive the full 15 points. Your instructor and TAs will NOT have access to your specific answers or overall score.

Those who complete BOTH online surveys will automatically be entered in a random drawing for an IPOD nano. One name will be drawn at the end of the semester and the winner will
be notified by email.

In addition, participation means you give your permission for researchers to have access to your course grade and certain coursework related to the study. All of this information will be kept strictly confidential.

**If you have any questions or concerns now or at any point you may contact:**
Kim Cossey, the principal investigator, by the contact information found on the handout.

Now I will hand out the paperwork.

If you wish to participate, check the box and sign and date the form. If you do NOT wish to participate, return the form UNSIGNED.
Section 3. IRB 21504 Sample Course Handout (Spring 07)

Extra Credit Opportunity for Chem 457

Spring 2007

Due dates for Extra Credit Surveys (Tuesday Sections)

January 24th - 29th (11:59PM) Presurvey

February 7th - 12th (11:59PM) Postsurvey

Surveys

Surveys can be found at: https://cbt.uts.psu.edu/

You will need your penn state access ID (the first six digits from your school email address, e.g. knc122) and password.

Procedures to be followed

If you agree to take part in this research, you will be asked to complete TWO online surveys during the semester. Each survey should take no more than 30-50 minutes each time, and will be worth 15 extra credit points. Surveys will ask questions on your perceptions of certain aspects of the course. Also, you will be asked to answer questions related to certain topics within the course. It is important to note that you will receive extra credit points based on the completeness of surveys, NOT on the correctness of your answers. Meaning if you complete a survey, you will receive the full 15 points. Your instructor and TAs will NOT have access to your specific answers or overall score. You have the right to not answer specific questions on the survey.

Those who complete BOTH online surveys will automatically be entered in a random drawing for an IPOD nano. One name will be drawn at the end of the semester and the winner will be notified by email.

If you agree to take part in this research, you are giving the researchers permission to make copies of your course assignments and test responses and use them for research purposes. This information will be used to evaluate the activities in the course. You are giving permission for your course grade to be used as part of the study.

Alternative

If you do not wish to participate in the surveys, you can alternatively receive an equal amount of extra credit by writing a 1 page review of an article from a scientific journal (email Kim Cossey, knc122@psu.edu for details). These article reviews would be due on the same dates as the corresponding extra credit activities, and be worth 15 points each.

NOTE: You cannot “double-up” extra credit by completing both an activity and a report in the same assigned period.

Contact Information

If you have any questions or concerns now or at any point you may contact:
Kim Cossey, the principal investigator, by phone at ###-#### to by email at knc122@psu.edu

Alternatively, you may contact Dr. Karl Mueller (ktm2@psu.edu), advisor to the project.
APPENDIX E

Depth of Understanding Rubric

Depth Q1
Question 1 In general, what is NMR? What does it do?

Count each category only once. The number of correct categories (and whether or not there are misconceptions) will determine the rating

Score sheet

<table>
<thead>
<tr>
<th>Data Set</th>
<th>ID</th>
<th>Depth Q1</th>
<th>magnetic</th>
<th>purpose</th>
<th>spins</th>
<th>process</th>
<th># C category</th>
<th>#M</th>
<th>depth1 rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>57S07</td>
<td>0.08182</td>
<td>Nuclear magn...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57S07</td>
<td>0.20734</td>
<td>It is a method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Categories: magnetic, purpose, spins, process - write # of correct responses in each box (see examples below)

# C category = sum of categories that contained correct responses, range from 0 - 4
# M = # of misconceptions, does not matter what category they are in

Correct Categories

A. Magnetic - example answers
   - Nuclear Magnetic Resonance = magnetic field = B field

B. Purpose - example answers
   - used to find structure = functional groups = number of H’s = chemical make-up
   - chemical information = electronic information = environment the nucleus is in
   - “characterize” alone is too vague

C. Spins - example answers
   - spin states
   - energy differences

D. Process - example answers
   - rf is applied = applied field at (90/180) degrees
   - relaxation occurs
   - converts data in time domain to frequency
   - Fourier transform is required

Ratings for Question 1

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>vague answer (no correct)</td>
<td>1 correct + M</td>
<td>2 correct + M</td>
<td>1 correct</td>
<td>2 correct</td>
<td>*3+ correct</td>
<td></td>
</tr>
<tr>
<td>all M</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n/r</td>
<td></td>
<td></td>
<td></td>
<td>3+ correct + M</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 3+ correct = 3 or MORE correct

Depth Q2
Question 2 There are many types of NMR experiments. What types of information about a
molecule/reaction can you get using the NMR spectrometer?

Equivalent answers are shown below with an ‘=’ sign. For example, the following response: NMR is used to show structure and functional groups scoring: receives credit for one correct answer b/c structure and functional groups are the same type of information.

For this question, misconceptions are NOT counted!

Ratings for Question 2

<table>
<thead>
<tr>
<th></th>
<th>0 correct</th>
<th>1 correct</th>
<th>2 correct</th>
<th>3 correct</th>
<th>4 or more correct</th>
</tr>
</thead>
</table>

Examples of Correct Answers

NOTE: Equivalent answers are marked with ‘=’

- structure = identity of molecule = number of H’s = bonding arrangement = functional groups
- used for different nuclei = used for C (or explicitly state other non-H nuclei, \(^{13}\)C, N, etc.)
- configuration = conformation = shape = arrangement in space = distance between nuclei
- H-C coupling = which H bonded to which C = bond connectivity C-H
- stereochemistry = diastereomers
- purity = percent yield = reaction completion = ratio starting materials to products
- ratios of molecules = relative quantities in a mixture
- what by-products = what impurities
- quantitative data = concentrations
- tracking and identifying intermediates = identify (or track) reaction intermediates
- track isotopic labeling = where certain isotopes are in a molecule
- relaxation constants (T1/T2) = effective correlation time (tc or tau-c)
- atomic motion = molecular rotation = bond rotation
- INTERmolecular interactions = diffusion = collisions = binding
- dynamics
- kinetic information about a reaction = rates of reaction

Incorrect (0 points):

- bond strength
- size = MW (very indirect)
- H bonding
- density
- 2D NMR (this answer is too vague)
- reaction information (this answer is too vague)
- NMR terms such as, frequency, spin, shift, coupling (hydrogen only coupling) etc. (not explain what this info tells about the MOLECULE)
APPENDIX F

Problem Solving Rubric

The problem solving measure only has one free-response item, the HETCOR problem. A sample taken from the scorer’s sheet for the HETCOR scorer’s sheet is shown below. When scoring items, please consider response 1 and response 2 as ONE answer!

A key for the columns on the scorer sheet is shown below. If the student response has misconceptions, the check the #M box. The other boxes are used to record correct information of the listed type. For example, if a student says “there are 5 different types of hydrogen” then you would put a check in the Th box.

<table>
<thead>
<tr>
<th># M</th>
<th>Connected? or Combining Info</th>
<th>Th</th>
<th>#H</th>
<th>Tc = #c</th>
<th>Sp</th>
<th>Fg</th>
<th>RATING</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of misconceptions</td>
<td>see accepted responses* below</td>
<td>unique types of hydrogens</td>
<td>number of hydrogens (integral)</td>
<td>types of carbon &amp; number of carbons/ same answer</td>
<td>splitting</td>
<td>functional groups</td>
<td>your rating</td>
</tr>
<tr>
<td>any #M (1 or more) will be counted the same</td>
<td>write “connected” or type of combining (“ester” or “alkene”)</td>
<td>mark as + or -</td>
<td>mark as + or -</td>
<td>mark as + or -</td>
<td>mark as + or -</td>
<td>mark as + or -; may be multiple answers, record all</td>
<td></td>
</tr>
<tr>
<td>CORRECT ANSWERS &gt;</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Please use the following rating scale, for scoring responses. Types of information are arranged into levels on the scale, however it is not necessary that every type be present in the response for that score to be received. For example, if a student correctly states that there are two methyl groups, but does NOT state how many carbons are present, she can still receive a rating of 4 (assuming she has stated no misconceptions).

**RATINGS**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Response</td>
<td>All M</td>
<td>Fg+H+M = #C +M</td>
<td>combining info (OR connected) + M</td>
<td>FG + #H</td>
<td>#C=Tc</td>
<td>combining info OR connected</td>
<td>combining info AND connection</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{M} = \text{misconceptions (1 or more)} \]

- 7 = combining info. AND connected - write which ones (connected, ester, and/or alkene)
- 6 = combining info. OR connected - write which ones (connected, ester, or alkene)
- 5 = number of unique C’s - this is not information they learned in Chem 36, inferring that (number of peaks) means the same as it would for H spectra
- 4 = correct functional groups, types/number of C, and/or types/number of H - chem. 36 level
- 3 = combining info. OR connected + \[ \text{M} \] - still getting something beyond Chem 36 level
- 2 = functional groups, types/number of C, and/or types/number of H + \[ \text{M} \]
- 1 = all misconceptions (all wrong answers) or nothing relevant, but they gave a response
- 0 = no response, didn’t try at all
APPENDIX G*

Research Methods for Proposed Program Evaluation Assessment (Chapter 5)

* Note: The following was modified from a course assignment for EdPsy 560, written by Kim Cossey, Catherine Goffreda, and Brian Hutchinson.

I. In-depth Interviews (IDIs)

In-depth, qualitative interviews are excellent tools to use in planning and evaluating programs. An in-depth interview is an open-ended, discovery-oriented method that is well suited for describing both program processes and outcomes from the perspective of the target audience or key stakeholder. The goal of the interview is to deeply explore the respondent’s point of view, feelings and perspectives. In this sense, in-depth interviews yield information. There are key characteristics that differentiate an in-depth, qualitative research interview from a regular interview. Some key characteristics of in-depth interviews include:

- Open-ended Questions. Questions should be worded so that respondents cannot simply answer yes or no, but must expound on the topic.
- Semi-structured Format. Although you should have some pre-planned questions to ask during the interview, you must also allow questions to flow naturally, based on information provided by the respondent. You should not insist upon asking specific questions in a specific order. In fact, the flow of the conversation dictates the questions asked and those omitted, as well as the order of the questions.
- Seek understanding and interpretation. You should try to interpret what you are hearing, as well as seek clarity and a deeper understanding from the respondent throughout the interview.
- Conversational. You should be conversational, but your role is primarily that of a listener. There should be smooth transitions from one topic to the next.
- Recording responses. The responses are recorded, typically with audiotape and written notes (i.e., field notes)
- Record observations. You observe and record non-verbal behaviors on the field notes as they occur.
- Record reflections. You record your views and feelings immediately after the interview as well.
In essence, in-depth interviews involve not only asking questions, but the systematic recording and documenting of responses coupled with intense probing for deeper meaning and understanding of the responses.

II. Observational Data Collection

Observational data collection provides the assessment leader the opportunity to document observable behaviors and results in the program’s natural environment without having to depend upon the participant self-report. For this evaluation, we will be using mostly structured observational methods to evaluate the fidelity of the program’s implementation to the program’s design. Limited unstructured observational data will be collected in the form of observed examples of specific behaviors to enrich and provide context for the observations. Data collection instruments will be comprised of observational checklists detailing specific behaviors one might expect from both teachers and students in the implementation of the new NMR curriculum at Penn State. Field note prompts will be embedded in the checklist for the collection of examples and unstructured comments. Observations will be conducted for two populations: teachers and students. The instruments will be designed so that the evaluation team will collect data on both populations during each classroom observation visit.

III. Focus Groups

Similar to in-depth interviews, focus groups provide an opportunity to gain perceptions and rich, detailed information on program and school functioning from key stakeholders. By developing a personal rapport, the interviewer can gain the perspectives and concerns of a common group of informants far beyond that of a survey or needs assessment. Each of the focus groups conducted for the program evaluation will include the following format and key characteristics:

- **Preparation.** Questions will be generated in a clear, straightforward fashion based on the respective goals for information gathered from each group. The interviewer will be selected on the basis of expertise in educational evaluation, speaking clarity, listening skills, and critical thinking abilities. Namecards will be provided for each participant, and every focus group will be audio or video recorded (pending authorization waivers from participants).
• **Question Construction.** Each question will be worded at the level of the respective participants. However, they will be comparable across stakeholder groups. They will be focused on the goal of the focus group, avoiding diversions and leading questions wherever possible.

• **Conducting the Group.** Ground rules, appropriate behaviors, and focus group objectives will be provided at the outset. Prepared questions will be followed systematically, with room for follow-up inquiries as needed. In the case of hesitancy to respond by participants, the interviewer will insert pauses of 5-10s to prompt thoughts. He or she will also provide only neutral responses to participants' comments.

If time and funding allow, follow-up groups with more diverse participant pools will be conducted to garner a range of perspectives and insights from each key stakeholder group.

**IV. Surveys**

Surveys are a popular method of evaluation due to their ease of dissemination, scoring and analysis. They key is to narrow focus through the choice of questions and wording, which must be precise to avoid inconsistent answers. For this reason, development is the most labor-intensive step of the process. Statistics can be performed on the data collected to provide evidence of a test’s reliability (and validity). Fortunately, it is often possible to obtain instruments from the educational literature (and test archives) that have been refined and tested for reliability and validity for a variety of student populations.

The goal of surveys is to gather many small pieces of information quickly and easily, often from a large pool of respondents. This will be accomplished by using multiple choice, ranking (Likert type) and (very few) open-ended items. The method of delivery of these items will vary based on the population.

• **Student surveys.** Responses will be gathered outside of normal class time using the online testing system at PSU, and items collecting different types of information will be combined into one instrument to reduce the number of interactions. Dates of data collection will be scheduled with the teaching faculty prior to the start of the semester when testing will take place.

• **Teacher surveys.** These will be distributed/collected via email. Upon sending, a date of return will be issued. Reminders will be sent (via email) as needed.
Students are arguably the most important population for this study and therefore testing of *all* participants would be optimal, therefore incentives will be offered. Surveys created by the evaluation team will be constructed to take as little time as possible while, will giving valuable information on implementation and effects of the program from the student perspective.

Surveys that pinpoint process information will be distributed/colllected via email to all teachers participating in the program. Reminders will be issued as needed.
Vita
Kimberly Cossey

Education

Ph.D., Chemistry (Chemical Education) August 08
The Pennsylvania State University, Advisor: Karl T. Mueller

M.S., Chemistry (Organic Synthesis) May 07
The Pennsylvania State University, Advisor: Raymond L. Funk

B.S., Chemistry (summa cum laude) May 00
Oklahoma Baptist University, Shawnee, OK

Teaching Experience*

*Took place at the Pennsylvania State University, University Park, PA unless otherwise indicated

NSF Fellow in GREATT Program, Aug 06 - Aug 08
K-12 outreach in Pennsylvania schools through the Graduate Research and Education in Advanced Transportation Technologies program, Department of Engineering

Guest Lecturer for Undergraduate Chemistry Courses Aug 03 - Aug 06
Courses included: Environmental Chemistry Lecture Course (~ 30 students), Organic Chemistry Lecture Course (~ 100 students), Honor’s Organic Chemistry Lecture Course (~20 students)

An Intern in Public Science Education (IPSE) Aug 04 – Aug 05
K-12 outreach in cooperation with the Franklin Science Museum and the Materials Research Science and Engineering Center (MRSEC) program

Teaching Assistant for Undergraduate Chemistry Courses Aug 02 - May 04
Organic and General Chemistry Lecture & Laboratory Courses (including honor’s courses)

Instructor for Undergraduate Course, May 00 – Aug 00
Department of Natural Science, Oklahoma Baptist University, Shawnee, OK

Workshop and Activity Development

2. Cossey, K. “Touchable Chemistry” GREATT Summer Teacher Workshop (Aug 07)
3. Kimball, A., Earnheart, K. “That’s So Gross” Franklin Science Museum (Philadelphia, PA), Homeschool Workshops (Nov 05)
4. Earnheart, K. “Microencapsulation” IPSE Homeschool Workshop Series (Apr 05) (http://www.ipse.psu.edu/activities/microcapsules/)

Additional Research Experience

Undergraduate Research (REU) for Oklahoma University May 99 - Aug 99
Synthesis of O-aromatic esters from amides at Oklahoma University, Norman, OK

Publications