COMPARISON OF HOMEWORK SYSTEMS (FOUR WEB-BASED) USED IN FIRST-SEMESTER GENERAL CHEMISTRY

Joshua Belland, B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

May 2009

APPROVED:

Diana Mason, Major Professor
Guido Verbeck, Committee Member
Michael G, Richmond, Chair of the Chemistry Department
Michael Monticino, Interim Dean of the Robert B. Toulouse School of Graduate Studies

Belland, Joshua, Comparison of Homework Systems (Four Web-Based) used in First-Semester General Chemistry, Master of Science (Chemistry), May 2009, 42 pp., 12 tables, 1 illustration, references, 7 titles.

Web-based homework systems are becoming more common in general chemistry as instructors face ever-increasing enrollment. Yet providing meaningful feedback on assignments remains of the utmost importance. Chemistry instructors consider completion of homework integral to students' success in chemistry, yet only a few studies have compared the use of Web-based systems to the traditional paper-and-pencil homework within general chemistry. This study compares the traditional homework system to four different Web-based systems. Data from eight, semester classes consisting of a diagnostic pre-test, final semester grades, and the number of successful and unsuccessful students are analyzed. Statistically significant results suggest a chemistry instructor should carefully consider options when selecting a homework system.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The affects of the World Wide Web on educational practices are undeniable. From online classes to online degree programs, there is an increasing use of the information distribution abilities of the Internet in the field of education. Students have come to expect class-specific web pages from their professors, and it is not unusual to have part of the class contain Web-based materials while students still attend classes in a traditional way (face-to-face with an instructor in a classroom), usually referred to as blended learning. In this study, Web-based homework was used as required supplemental work giving the students practice with instructor assigned homework problems.

A query of Google for "online chemistry homework" returns over 300,000 hits. A vast majority of these results are probably for chemistry homework tutoring sites (one might get the idea that students often need a lot of additional help in understanding how to complete chemistry homework problems), but within these hits are results for chemistry homework software systems that utilize the Internet for the distribution, collection, and grading of chemistry homework problems. Quickly located in these results are programs called: MasteringChemistry, OWL (Online Web-based Learning), WebAssign, and several others.

With a steady increase of available software systems for chemistry teachers to use for assigning homework, two questions must be asked:

- Are online chemistry homework systems significantly different than the traditional system of homework problems from a book or worksheet in positively affecting student's success in chemistry class?
- Is there any significant difference in students' success with any one of the many chemistry homework software systems available compared to the others?

Fortunately, a scenario for deriving answers to these questions presented itself in the form of Dr. Mason's General Chemistry for Science Majors I class at the University of North Texas. Student performance data are available from the last four years (eight fall and spring semesters) allowing for statistical comparisons between using traditional textbook homework problems versus various online homework systems. I will utilize these data to quantitatively derive answers to the two questions stated earlier.

At the onset of this research, it is my educated guess that there will be no statistically significant differences in the performances of any of the semester classes, to the alpha level of 0.05. The null hypotheses that will be tested are:

1. There is no statistically significant difference between the classes that did not use an online homework system and those that did.
2. There is no statistically significant difference between any of the classes that used any of the four online homework systems.

It is exceptional to find studies that compare the use of Web-based homework to paper and pencil homework within the field of chemistry. It is even more difficult to find one that is able to compare several Web-based homework systems to other Web-based systems - I could not discover any such research available with ERIC. Thus my study is very unique in design, and the implications will be of interest to instructors of chemistry, to textbook authors and software designers, and of course to students.

As all chemistry instructors have to cope with ever-increasing workloads as more students take chemistry classes, and desire a way for students to have viable practice over chemistry concepts with relevant feedback. Completion of homework is considered essential to a student's success in chemistry, yet students tend to be unmotivated in completing homework if there is no feedback system to acknowledge correct answers and offer remediation for incorrect answers. This study will help illuminate if a Web-based homework system could be not only a benefit to a chemistry instructor but also to the students.

Furthermore, the end result could allow for a greater understanding of what tools will help a greater number of chemistry students be successful in chemistry. Successful completion of freshman chemistry courses is central to obtaining many post-secondary degrees, mostly in the sciences and engineering but also in disciplines such as psychology, nursing, some fine arts, etc. As the world culture shifts to place a greater emphasis on the study of the sciences, the role of chemistry is evident. Even at the high school level, now that Texas has adopted the 4 X 4 curriculum, chemistry is usually a requirement for graduation of high school.

Before continuing, there will be several specialized terms used throughout the following chapters that should be clarified:

First-Semester General Chemistry: General Chemistry for Science Majors I offered at the University of North Texas and taught by Diana Mason, PhD

Successful students: University of North Texas students receiving final grades in General Chemistry for Science Majors I of A, B, or C

Unsuccessful students: University of North Texas students receiving final grades in General Chemistry for Science Majors I of D, F, or W

Supplemental instruction: instructional tasks given outside of the scheduled class time, referring to homework problems in this study

Web-based instruction: using the Internet to store, disseminate, and record information in the form of graphics, questions, responses, and other multi-media

Homework system: method of presenting homework, either as assigned questions from the chapter of the course textbook or assigned questions presented to the student via the Internet

TAMS: Texas Academy of Mathematics and Science, a program for high school age students that allows them dual credit (both at high school and at the University of North Texas) for classes taken at the University of North Texas

CHAPTER 2

REVIEW OF THE LITERATURE

A search with ERIC for literature on comparing the use of different Web-based homework systems reveals no other studies of this topic; evidence that this study is unique within the subject of chemistry. Broadening the search criteria results in successfully finding literature about the use of a single, Web-based homework system in a course, but it is often not a chemistry, or even science, course. Examination of these articles has led me to believe that the results found in the literature mentioned herein should be applicable to particular points of my research study.

In general the following trends have been noticed in the literature and are applicable to my study:

1. Most studies show little or no measurable gains in student performance with Web-based homework compared to paper and pencil homework. This is significant in that no study was uncovered that presented a case for a detrimental effect of using Web-based homework.
2. If any significant difference in scores were found, then an increase in scores with Web-based homework is only found for the otherwise lower achieving students (those that would otherwise be unsuccessful) with no significant difference in the scores of otherwise higher achieving students.

3. Students will spend more time on the homework if it is graded than if it is ungraded. Instructors feel this "time on task" is a worthwhile expenditure of the student's time for success in the course.
4. All articles mentioned the positive effects of using Web-based homework for the instructor of the course: From the time savings of not having to grade homework, to the ability to accommodate an increased enrollment in classes, to the absence of large amounts of students seeking help during office hours.

To put figurehead to the fourth trend above, an article[1] by Sanders, published in 2005, describes the process and results of adding a significant Web-based portion to what was otherwise a traditional Spanish courses. Sanders boldly states in the abstract that the redesign of the Spanish courses to include Web-based components allowed for an enrollment increase of 85%, a decrease in the cost-per student to the department by 29%, and subsequently a monetary raise to the course instructors of 9%.

The article goes on to explain the process of installing the Web-based components and gives convincing justification to the claims stated in the abstract. Unfortunately, there is no reason to believe that the results for Sanders' Spanish study can be generalized to a chemistry course. Still it should cause one to consider that there might be benefits to using Web-based homework that are not seen in just the students' scores.

Furthermore, while summing up the advantages of Web-based components in the Spanish courses, Sander estimated the instructors were saved 90 minutes per section per week that were previously spent on grading homework. This result is not isolated to this particular study, but is typical of many of the applicable reports on the use of Web-based homework. Although the actual estimate of 90 minutes per week might change from course to course or instructor to instructor, few would doubt that this time might be better spent with increased office hours to help students on an individual basis or in some other fashion that allows an instructor meaningful one-on-one interaction with students.

Cole and Todd of the University of Wisconsin-Madison provide the most similar research to my study with their work on the "Effects of Web-based Multimedia Homework with Immediate Rich Feedback on Student Learning in General Chemistry"[2], published in 2003. Cole and Todd used WebCT (a system for delivering classroom content online, used by many colleges and universities) in attempt to improve student performance for students in a fall, first semester general chemistry class. They saw the use of WebCT delivered homework (or quizzes - an explanation will be found in the discussion of the experimental procedures) to be a means of reaching three goals for the students in the class:

- Help students to identify their own misconceptions and areas needing help, without direct intervention from an instructor.
- Provide immediate feedback on homework (or quizzes).
- Be able to use animations and multimedia for illustrations of chemical phenomenon.

These goals constitute a good representation of most chemistry instructor's intentions when choosing to use a Web-based homework system, thus the purposes of their experiment should be easy to generalize to the interests of the population of chemistry instructors as a whole.

As typical as I feel their intentions to be, their experimental design appeared rather peculiar. All the classes in the experiment were divided into even course number groups and odd course number groups. The odd number group were assigned the WebCT Web-based homework while the even number students were assigned traditional paper and pencil homework. Cole and Todd defend this choice in that students were just as likely to be registered into one class as another by the university's registrar system and each teaching assistant was assigned to an odd and an even number class to be able to compare any effect of individual teacher assistant attitudes on the experiment.

Students in the control (paper and pencil homework) group were assigned textbook questions each week. Teaching assistants graded (after a little training) the weekly homework assignments for completion, not accuracy (correctness). Homework questions were graded on a scale of 0-3 points as the teaching assistants scanned the assignments. Thus students that seemed to spend a lot of effort solving a problem on paper, but provided an incorrect answer might receive more points than a student that did only a little work on the paper, but through a significant mental effort could provide the correct answer.

The experimental (WebCT) group was assigned Web-based homework, delivered through WebCT's quizzing capabilities. Students logged-in and accessed weekly homework assignments consisting of questions randomly selected by WebCT from an instructor created pool of questions. As a result, most students would have only a few questions in common with anyone else in the entire experimental population, and it would be statistically very unlikely that any two students would have exactly the same set of questions. This allowed for students to consult with each other in attempt to get the right answers without the chance of a student being able to simply copy the correct answer from another student.

Students in the experimental group did receive a grade for the homework assignments based on correctness. WebCT's quiz ability graded the assignment immediately after the homework was submitted, and with incorrect responses gave feedback directing the student to textbook sections, online course material, or auxiliary course materials available at campus computing centers. Students were allowed to take the homework quiz twice per weekly assignment; with the best score kept (the likelihood of getting the same question twice within the week is low as the second attempt was again comprised of randomly chosen questions from the question pool). Students were also encouraged to make their first attempt early in the week to be shown the deficits in their chemistry understanding and then take the second attempt later in the week after hopefully understanding what they were missing.

In light of the very different conditions between the two groups, a thorough quantitative and qualitative analysis was summarized by Cole and Todd. The overall results of the study showed that Web-based homework could *not* be considered any more effective in increasing student scores than paper and pencil homework. Only if the points for homework were added into the student's overall scores was there a statistically significant difference between the experimental and control groups (the WebCT students tended to get higher homework scores, most likely because of the feedback system). This supports trend number 1.

It was noted in the qualitative findings the students with a higher GALT score (Group Assessment of Logical Thinking; considered a logical thinking aptitude test) suggested they would prefer to have paper and pencil homework whereas the students with a lower GALT score stated they would rather use the Web-Based homework system, in support of trend number 2.

In support of trend number 4, Cole and Todd write "[t]he potential for personalized, detailed, rich feedback to the students at low cost to the instructor in terms of time spent grading is an advantage that should not be overlooked." (p 1342)

Of all the previous studies that will be examined herein, this study shows the most potential in being applicable to the population of first semester general chemistry students that are the subject of this study, so the results should be seen as a good indicator of what will probably be seen in my study.

Similar studies to that of Cole and Todd's could not be found, so at the risk of losing the ability to generalize to the sample population of students in my study, I broadened my search criteria and found a study involving organic chemistry students and a Web-based homework and quiz system called WE_LEARN. Penn, Nedeff, and Gozdzik published a study in 2000 entitled "Organic Chemistry and the Internet: A Web-Based Approach to Homework and Testing Using the WE_LEARN System"[3].

WE_LEARN is an acronym for Web-based Enhanced Learning Evaluation and Resource Network and was created to meet five design specifications:

- Provide the students and teachers with immediate feedback
- Provide an active learning environment

- Provide similar and dissimilar learning environments
- Provide an enhanced learning system
- Provide a high level of flexibility

Having met these criteria with WE_LEARN, Penn, Nedeff, and Gozdzik reported test scores are higher when comparing the semesters using WE_LEARN to the semesters using no online system. Unfortunately, no statistical analysis was reported in this article, but, as already mentioned, it is doubtful that the population of students found in the organic chemistry classes would be the same population found in the general chemistry classes involved in my study. Therefore, assuming the worst – that the changes in test scores are due to random chance – the fact that all test score changes have coincidentally been upwards should still logically support the first trend stated at the beginning of this chapter.

More impressive is the evidence Penn, Nedeff, and Gozdzik cited that supports trends 3 and 4. An advantage of WE_LEARN was the ability to track the number of students using the system and the times they use it. Not surprising, students were increasingly likely to log in and take a review quiz shortly before a class exam, with increasing numbers of students taking quizzes as the semester progressed. Apparently, students were willing to spend more time outside of class on the subject as they recognized their test grades could be subsequently improved with more practice, similar to the third trend.

Another, unexpected and exciting, advantage was the gains in time made by the course instructor due to the change in the students' understanding (if I may be so bold as to call it that), when comparing student behavior pre- to post- WE_LEARN. Penn, Nedeff, and Gozdzik report before using WE_LEARN an estimated 20-30 students on average visited the instructor's office per exam week, requiring an estimated 20 hours to answer their questions (it is not reported if this time was solely put in by the instructor or if there were additional people present to help answer questions). After implementing WE_LEARN, Penn, Nedeff, and Gozdzik report an actual average of three students per exam week, which only required an average of 15 minutes to answer their questions.

Moreover, it would seem the students who were still seeking help were asking very different questions. As Penn, Nedeff, and Gozdzik state it, "Before the implementation of the system, the majority of questions posed by students … showed almost no comprehension of the basic concepts. After the implementation of WE_LEARN, this type of question disappeared almost completely. In fact, the routine types of questions … were very specific to the types of reaction conditions or subtle structural variations that might lead to a different set of products." (Penn, Nedeff, and Gozdzik, p 230)

Additionally the attendance at help sessions held the evening before the day of an exam dropped dramatically from an estimated 75% of the class enrollment pre-WE_LEARN to 33% of the class enrollment after implementing WE_LEARN. All these effects added together would equal to an enormous time advantage for the instructor and any supporting staff, affirming trend number 4.

Penn, Nedeff, and Gozdzik's study should not be seen as specifically applicable to my study, but does reinforce the general trends found throughout the literature. If the target population is broadened even more to include all sciences, then there is a noteworthy study involving some Introduction to Astronomy classes.

Allain and Williams' article "The Effectiveness of Online Homework in an Introductory Science Class"[4] from 2006 was chosen because of its experimental design was very intentional in trying to compare the use of Web-based homework to traditional paper and pencil homework. Their results are typical, but the design of the course did allow for statistical analysis to be performed on the results.

Although Allain and Williams' students in four sections of Introduction to Astronomy are non-science majors, and completely outside of the sample population involved in my study, I feel the intentional design of how each section of the course did homework, the application of a pre-test and post-test to measure conceptual understanding, and the comparison by statistical analysis allows for the results of their study to be applied (with reservations) to the results that I expect to see with my study's sample of chemistry students.

Allain and Williams used a software program called WebAssign as a way of assigning and grading homework problems via the Internet. The Introduction to Astronomy course had four sections of about 40-90 students per section, and for the sake of ease, each section was chosen as to have a different form, or combination, of homework. One section was assigned homework using WebAssign for the entire semester. Another section was assigned homework as paper and pencil exercises from the textbook, but these were not collected or graded. A third section spent the first half of the semester using WebAssign, and the second half of the semester with paper and pencil homework (with an additional research paper). The last section was the inverse of the third, with the first half of the semester involving paper and pencil assignments (and the research paper) and the second half of the semester with WebAssign homework.

Allain and Williams administered the pre- and post- forms of the Astronomy Diagnostics Test as a way to measure changes in student's conceptual understanding. Allian and Williams used the test results to calculate a normalized matched gain (prescore less the postscore, divided by the maximum score less the prescore) for each student. Once statistical analysis of the different sections and diagnostics tests were analyzed, Allain and Williams conclude that there was no significant difference in conceptual understanding or test scores between any of the sections. They did find that students report spending more time on the homework when it was graded (WebAssign) than otherwise (paper and pencil), but this additional time-on-task did not increase test scores, and by inference, the students' understanding.

Two things are apparent in this study. First, the results match trend 1. Second, as homework was only graded when using WebAssign (the grading would be completed automatically) I would infer it also supports trend 4, even though it is not explicitly stated in the article. Obviously it was a foregone conclusion that grading paper and pencil homework would take a considerable amount of time, more than the instructors were willing to spend.

Additional support for trend 1 can be found with several other studies that will be mentioned to show that the trends are not limited to chemistry or even the subject of science, but are rather common trends for the use of a Web-based homework system, regardless of subject.

In the 2007 article by Zerr "A Quantitative and Qualitative Analysis of the Effectiveness of Online Homework in First Semester Calculus"[5] a Web-based homework system with an attempt-feedback-reattempt feature was compared to paper and pencil homework (the paper and pencil homework was not graded). Zerr reports no statistically significant differences for an alpha level of 0.05, although the qualitative reports indicated the online homework might have helped those that otherwise would have received D's and F's to achieve passing grades (this article is the strongest voice backing trend 2).

In the 2005 article by Hauk and Segalla "Student Perceptions of the Web-Based Homework Program WeBWorK in Moderate Enrollment College Algebra Classes"[6] homework delivered with WeBWorK was compared to paper and pencil homework assignments (which were also graded). A pre-test and post-test system was used to check for achievement differences between students assigned to WeBWorK or to paper and pencil homework, and no statistically significant differences were discovered. All the instructors involved with classes using the WeBWorK homework assignments considered it a timesaver.

Additionally, 78% of the students using WeBWorK made efforts to do the homework, compared to 65% of the students using paper and pencil homework. Psychologically speaking, the students that turn in homework tend to be successful (make a C or higher) and the students that don't turn in homework tend to be unsuccessful, so it might be inferred that 13% of the students in the WeBWorK group would have otherwise been unsuccessful students that with the use of Web-based homework may have been successful in passing the course. This is pure conjecture, Hauk and Segalla do not make any claims regarding this and did not include data in the article that could be used to test this thought, but I feel it's not unreasonable to consider that at least a part of the WeBWorK students would otherwise have received lower grades, fitting trend 2.

With this overview of the applicable literature, I expect it's apparent that I would not be surprised to see the same four trends continue in my study, even with the additional component of using many different Web-based homework systems. I am not expecting to see a statistically significant difference in grades between any of the semester classes.

CHAPTER 3

STUDY DESIGN

As mentioned earlier, the subjects involved in my study are students taking a first semester of General Chemistry for Science Majors class at the University of North Texas, specifically having Dr. Mason as the instructor of the course. Some of the students in these classes are atypical college students, and this will need to be accounted for before a genuine comparison can be performed. Data from eight semesters will be used, from the fall of 2004 through the fall of 2008. Two of these semester classes used paper and pencil homework and the other six classes used a third party Web-based homework system.

To keep the data manageable, the main comparison of the classes will be by final semester means. Additionally, all the classes took a diagnostic pre-test that will allow for a comparison between the classes and to the entire population of college/university students in the United States that also take this test.

Students enrolled in general chemistry I classes are diverse as one usually considers – with different educational, racial, and ethnic backgrounds, but this is not the diversity that will necessitate consideration during analysis, as the relative percentages of each demographic are reasonably similar each semester. The atypical students referred to are a part of the TAMS program.

The Texas Academy of Mathematics and Science (TAMS) program is a residency program for high school aged students as they attend university classes at the University of North Texas during what would otherwise be their junior and senior years in high school. Four of the eight semesters considered in this study include TAMS students taking general chemistry I in the fall. These students, in some semesters, comprise up to 50% of the students enrolled in the class and certainly cannot be considered as from the same population as a typical college student for three reasons:

First, the TAMS students are of high school age, yet are taking college courses. It is easy to infer that these are extremely bright, gifted, and hard working students. In all four semesters under consideration, the mean semester grade of the classes drop when the scores of the TAMS students are removed (these data will be shown later). I feel it would be unwise to consider TAMS students along with the typical UNT college student as we can infer that the TAMS students are probably better prepared academically, or at the least more likely to be successful on homework and tests with or without any type of intervention.

Second, the TAMS students have a built in support team to encourage their high performance. The adult administrators of the TAMS program monitor the students' performance in their classes and offer remedies if performance decreases. This monitoring certainly does not happen with the majority of the typical college students, who are often on their own in guessing what their class grade is if their professor doesn't post them, and are on their own to seek the necessary help to better their performance.

Third, students in the TAMS program must apply and be accepted to the program and therefore can also be dismissed from the program for poor performance (grade other than A or B or perhaps a limited number of C's). For a typical college student there is little threat of being expelled, as long as their average for all their classes is still above the designated limit set by the university.

Although IRB approval and parental permission was ascertained, it will be appropriate to remove the TAMS data from the final analysis. This will be the main way that data across the semester classes will be analyzed, although to verify I am justified in removing their scores I will begin the data analysis with the TAMS students scores retained.

Another issue that needs to be considered is the differences in students between the fall semester and spring semester classes. Dr. Mason's General Chemistry I course is designed to fit into the fall semester of a science major's freshman year. In general then, students that are taking the course in the fall are "on track" and students that do not take the class until the spring are probably having to take other remediation classes during the fall semester, or are students who have previously been unsuccessful in general chemistry I, often called "repeaters". Additionally, Dr. Mason has informed me that it is not uncommon for a typical UNT student to postpone general chemistry I until the spring for the purpose of not being in the same class as TAMS students (as if the presence of the TAMS students was intimidating).

Further differences exist between the fall semester classes and the spring semester classes that admittedly could limit the applicability of this study. In the fall the UNT offers more classes of chemistry than in the spring, so students taking the course in the fall have a greater choice of instructor and time of the day the class is offered than in the spring. Although Dr. Mason's classes are at the same time both semesters, it is possible that students in the fall that would otherwise be unsuccessful would choose a different instructor; one with a reputation of using an easier grading system. Students in the spring would have fewer choices of instructors, perhaps causing there to be a larger population of potentially unsuccessful students in Dr. Mason's class in the spring compared to the fall.

Also, a repeating student may not be successful even on their second attempt at the course, thus counting towards the measure of unsuccessful students twice. If the student was in Dr. Mason's class each time, then this student's data for their second attempt could simply be removed. However, unsuccessful students often try a different instructor's class for their second attempt, and unless these students would voluntarily report their past unsuccessfulness in another instructor's course it would be a violation of the students' confidentiality to determine they were unsuccessful in a previous class with a different instructor.

As this study is performed on data that has been mostly collected after-the-fact, it is not possible to add in a survey on which students might volunteer their information. Future studies on this topic would wish to include such a survey so an analysis of the effect of repeating students could be performed. I am unable to introduce such a survey so I will be unable to perform such an analysis. For this study each semester was treated individually, so a student that is repeating the course will be factored into the analyses the same as a student taking the class for the first time.

In the classes studied, the final semester scores consist of grades from four exams and homework grades. Unfortunately, I could not propose controlling the material found on each exam, nor other conditions of the classes (such as homework topics). However, after consulting with Dr. Mason, I feel confident enough in the similarities of each semester to continue this study, as the exams and most homework grades had similar placement in the course sequence (thus similar topics) each semester.

All classes took, as a pre-test, the California Chemistry Diagnostic Exam (CA Dx) distributed through the American Chemical Society Division of Chemical Education Examinations Institute. I will be using the diagnostic exam scores primarily as a way to compare the initial chemistry understanding of the students between each class, and secondly to evaluate how well the students in my study match the population of the students across the United States that also participated in this exam (which should be a valid estimate of the population of chemistry students in the United States).

The CA Dx was designed by a panel of sixteen post-secondary chemistry instructors in an attempt to provide a useful entrance exam that could be used as a predictor of students' chemistry success. In the article "A Rationally Designed General Chemistry Diagnostic Test"[7] from 1994, author Russell recounts the experiences of being on the development panel and reviews the statistics that were used to arrive at the present form of the exam.

The first version of the CA Dx was available in 1989 with a revised exam available in 1993, and for this study the updated 1997 version was used for all classes. The exam has 44 multiple choice questions allowing student answers to be quickly scanned, graded, and recorded.

Furthermore, Russell states clearly in the article that although the exam was a "good" predictor of student success in freshman chemistry courses, the best use, in her opinion, was to use the results to counsel students concerning their weaknesses and suggest methods to improve these before classroom performance was impacted. Russell discouraged the use of the CA Dx to exclude students from a chemistry course.

For the purposes of my study, the pre-test CA Dx scores will be limited to comparing my sample populations (students in Dr. Mason's course) to each other by class, and to the population scores kept by the ACS. With the very large sample sizes accessible from the ACS data, I feel confident that statistically it will be satisfactory to assume it equivalent to the entire population of students taking first-semester chemistry.

The final class means and CA Dx scores will allow for a comparison between the eight semester classes, promoting an analysis of the five homework systems. As a final point of comparison, the number of students that are successful and unsuccessful each semester will be compared. The rational is: the better the homework system - the greater the number of students that will be successful.

The first two semester classes (fall of 2004 and spring of 2005) were assigned homework as pencil and paper questions coming from a variety of sources; all were presented in printed form and answered by the students in printed form, with the grades included in the students' final scores.

In the next three semesters (fall of 2005, spring of 2006, and fall of 2006) the online homework system Online Web-based Learning (OWL) was used, and homework assignments were presented and completed using this system, with the homework grades included in the students' final scores.

In the fall of 2007 the online homework system MasteringChemistry was used. In the spring of 2008 the online homework system CATALYST was used, and in the fall of 2008 the online system SmartWork was used, with the homework grades included in the students' final grades for all of these semesters.

OWL is an online homework system by Brooks and Cole, now part of Cengage Learning at the University of Massachusetts in Amherst, written by Day, Botch, Hixson, Lillya, and Vining. Students receive an access code with the purchase of their textbook or may purchase an access code without purchasing the book. (Due to an agreement between Dr. Mason and the various publishers, none of the students enrolled were required to purchase the access codes outside the very first semester that OWL was employed.) Questions are categorized by different textbooks to match the specific topics covered in that textbook's chapters, and the instructor can decide which set(s) of questions to be delivered to the students.

Students answer fill-in-the-blank style questions, and if incorrect answers are provided by the student, then a rational for the correct answer is provided by the program. If correct answers are submitted, then reinforcing information affirming the correctness of the answer is provided by the program. Questions may be presented with diagrams or other graphical representations, whether specifically for students to use in answering the question or to remind students of the general topic covered by that question set. All students get questions from the same question set, but do not necessarily get the same questions, and before answering the question sets students get a chance to review the textbook chapter online, and sometimes also additional tutorial material by the writers. Questions are immediately scored and appropriate feedback provided.

Although OWL has seen significant upgrades after the semesters it was used for my study, the basic composition and methods of the program are the same; therefore I feel it is still appropriate to continue to use these data for these semesters in my analysis.

After the semesters that used OWL, the Web-based homework system MasteringChemistry (also known as Mastering General Chemistry) by Pearson was used. MasteringChemistry consists of a large question bank that correlates the questions to the specific chapter of the specific textbook chosen by the instructor. Access codes can also be purchased without the purchase of a textbook.

Large amounts of questions for each question type are available, so students could be given a randomized set of questions over the same topic and thus not have identical assignments. A database is kept for each question type that shows the average question scores, average time for completion, difficulty level, and other additional information that might be helpful when designing a homework assignment. This database is for all students that have ever completed that type of question, so over the years the sample population comprising these statistics had become extremely large, allowing the instructor to also gauge how students are performing compared to all the present and past students using this program. Graphics can be shown with questions, and students are able to enter formulaic answers if that is what the question call for. Feedback is provided in a feedback section of the screen for both incorrect and correct answers. Students may also access electronic versions of the textbook if they need to consult the chapter material for review.

MasteringChemistry has a built in tutorial option that the instructor can activate for students that guides the students through a problem in a Socratic fashion. The instructor can also set whether accessing the tutorial feature lowers the students' grade on that question or if students that do not access the tutorial feature get extra credit points. For this study it should be noted that the suggested protocol provided by the publishers was used (e.g., the students could get bonus points if the tutorial feature was not used). As before, assignments are immediately scored.

The following semester CATALYST by Wiley Publishing was the Web-based homework system employed. Access codes can, again, be purchased with the textbook selected by the instructor or can be purchased without a textbook. Questions for homework assignments can be selected in two ways: individual questions can be picked by the instructor with algorithmic changes to questions involving calculations so students do not get exactly the same question, or a series of questions concerning a specific topic can be picked, giving students a randomized set of questions; making each student's assignment unique. For this study it should be acknowledged that the question sets were selected for each unit. These sets consisted of about 15 questions each.

Questions not only can include graphics but there is a formula bar that allows for the entry of answers involving chemical formulas or chemical equations. Feedback is provided for reinforcement with incorrect and correct answers. Furthermore, there are a variety of help features. Questions often will have a "link to text" feature that opens the online textbook to the section covering the topic of the question. Some questions have a tutorial question option that students can access at any time while attempting the question that shows a similar-topic question worked out from beginning to end. Finally there is a hint option that could provide a starting place for a student that doesn't know where to begin. Again, assignments are immediately scored.

In the final semester of this study, the Web-based homework system SmartWork was used. Students purchase access codes along with the textbook selected by the instructor. SmartWork allows instructors to choose from ready-made assignments, individually pick questions, or create new questions to be used in the class system. Questions involving calculations have algorithmically changed data so students get numerically different questions.

Question types involve those that require word answers, formula and equation answers via an equation editor, and even structural answers using a grid that allows for Lewis structures and orbital notion to be submitted. As students work to complete the assignments, SmartWork offers feedback for incorrect answers, a hint option, example problems with solutions, and even a tutorials system. Also students can be directed to the corresponding section of the Norton E-book version of the textbook chosen for the course. Instructors can regulate if there are point deductions for accessing the help features. As with the other homework systems, assignments are immediately scored.

The following table might be helpful in summarizing the features of these four Web-based systems, but should be considered only a partial comparison. It is interesting how many features seem to be in common, but each system has its own unique interface and setup.

Table 1: Summary of the Homework System Features

Feature	OWL	MasteringChemistry	CATALYST	SmartWork
Questions matched with a textbook	X	X	X	X
Answers could contain formulas		X	X	X
Guiding responses (feedback)	X	X	X	X
Diagrams	X	X	X	X
Students may get different assignment questions/data	X	X	X	X
Links to an Web-based textbook	X	X	X	X
Homework scores are immediately available	X	X	X	X

CHAPTER 4

DATA ANALYSIS AND RESULTS

The best place to begin comparing classes is their mean semester course grades. In the chart below is listed the class by semester and year, the number of students in the class, the name of homework system used, the mean class grade for that semester, and the standard deviation. Note Dr. Mason did not teach a Spring 2007 general chemistry class.

Table 2: Mean Semester Grade by Class

Class	Number of Students	Homework System	Mean Semester Grade (SD)
Fall 2004	66	Paper and Pencil	79.2 (19.1)
Spring 2005	114	Paper and Pencil	65.2 (24.0)
Fall 2005	96	OWL	83.3 (17.7)
Spring 2006	148	OWL	67.5 (19.3)
Fall 2006*	94	OWL	74.8 (18.8)
Fall 2007	144	MasteringChemistry	83.7 (15.0)
Spring 2008	146	CATALYST	69.9 (19.5)
Fall 2008	116	SmartWork	79.5 (19.9)

*No TAMS enrolled.

As discussed earlier, for a correct comparison the TAMS students' grades need to be removed from the results for the fall semesters, as they are not representative of the typical chemistry student enrolled at UNT. To support this decision, a one-way ANOVA was performed on all eight classes with the scores from the TAMS students retained.

Table 3: ANOVA Results of Mean Semester Grade with TAMS Students Retained

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	41483.716	7	5926.245	16.218	.000
Within Groups	296346.33	811	365.409		
Total	337830.04	181			

Statistically significant results were obtained between the following semesters (with redundancies retained):

- Fall 2004 with Spring 2005 and Spring 2006
- Spring 2005 with Fall 2004, Fall 2005, *Fall 2006*, Fall 2007, and Fall 2008
- Fall 2005 with Spring 2005, Spring 2006, and Spring 2008
- Spring 2006 with Fall 2004, Fall 2005, Fall 2007, and Fall 2008
- *Fall 2006* with Spring 2005 and Fall 2007
- Fall 2007 with Spring 2005, Spring 2006, *Fall 2006*, and Spring 2008
- Spring 2008 with Fall 2005, Fall 2007, and Fall 2008
- Fall 2008 with Spring 2005, Spring 2006, and Spring 2008

It is easy to generalize that fall classes had statistically significant differences from spring classes, and vice versa. As TAMS students are only involved in the fall classes, it is evident the TAMS students' grades inflate the class mean.

It should be noted that the Fall 2006 class (shown in italics) was a fall class without any TAMS students enrolled. This is the only exception to the general trend just deduced. For the moment I suggest the results for the Fall 2006 class be ignored.

Therefore a proper comparison of the classes needs to be performed with the TAMS students' scores removed. Below is a comparison of the classes with the TAMS students' grades removed.

Table 4: Mean Semester Grade by Class without TAMS Students

Class	Number of Students	Homework System	Mean Semester Grade (SD)
Fall 2004	50	Paper and Pencil	73.8 (18.8)
Spring 2005	114	Paper and Pencil	65.2 (24.0)
Fall 2005	47	OWL	74.3 (21.0)
Spring 2006	148	OWL	67.5 (19.3)
Fall 2006	94	OWL	74.8 (18.8)
Fall 2007	71	MasteringChemistry	75.0 (16.2)
Spring 2008	146	CATALYST	69.9 (19.5)
Fall 2008	68	SmartWork	68.8 (21.0)

This time a one-way ANOVA yields only statistically significant differences between the Spring 2005 class and the Fall 2006 and Fall 2007 classes (there is not a statically significant difference between the two fall courses). Thus, the Spring 2005 mean class grade was significantly lower than the two highest classes' mean grades in the comparison (but only for the two highest).

Table 5: ANOVA Results of Mean Semester Grade without TAMS Students

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9300.736	7	1328.677	3.361	.002
Within Groups	252234.37	638	395.352		
Total	261535.11	345			

To check if the statistically significant differences were inherent in the classes by chance of enrollment, the mean class scores on the CA Dx were compared. This test was given within the first two weeks of each class and consisted of 44 questions (the highest score possible is a 44). The scores of TAMS students have been omitted.

Table 6: Mean CA Dx Score by Class without TAMS Students

Class	Number of Students	Homework System	Mean CA Dx Score (SD)
Fall 2004	45	Paper and Pencil	19.3 (6.8)
Spring 2005	101	Paper and Pencil	18.2 (6.0)
Fall 2005	38	OWL	19.5 (7.0)
Spring 2006	133	OWL	18.5 (6.3)
Fall 2006	76	OWL	19.7 (7.5)
Fall 2007	62	MasteringChemistry	19.1 (6.9)
Spring 2008	106	CATALYST	18.5 (6.2)
Fall 2008	48	SmartWork	22.1 (7.5)

Another one-way ANOVA yields results that can probably be guessed at just by looking at the data; the only statistically significant differences are between the Fall 2008 class and the three spring classes.

Table 7: ANOVA Results of Mean CA Dx Score without TAMS Students

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	636.552	7	90.936	2.058	.046
Within Groups	26551.123	601	44.178		
Total	27187.675	608			

The Fall 2008 students scored significantly higher on the CA Dx exam than usual, but it should be noted that the classes having statistically significant differences in their mean semester grades did *not* show statistically significant differences in CA Dx exam. Also note that although the Fall 2008 students start with apparently more chemistry understanding, by the end of the semester the final class's mean was not significantly different compared with other classes.

The Examinations Institute (part of the American Chemical Society) keeps normalized data on the scores reported back by the instructors that use the California Diagnostic exam. Comparing the mean scores above to the normalized score reported by the Examinations Institute of 20.45, with a standard deviation of 7.56 (since 1997) indicates the sample populations used in this study perform below the average of the entire population. However, the sample populations in this study are performing consistently between the classes, except in the one case noted.

The final comparison of the classes will be with the number of successful students and unsuccessful students in each class. The best way to view the data is by percentage as the total number of students each semester does vary considerably.

Table 8: Successful and Unsuccessful Students by Class without TAMS Students

Class	Number of Students	Homework System	Successful (A, B, C)	Unsuccessful (D, F, W)
Fall 2004	50	Paper and Pencil	30 (60%)	20 (40%)
Spring 2005	114	Paper and Pencil	79 (69%)	35 (31%)
Fall 2005	47	OWL	25 (53%)	22 (47%)
Spring 2006	148	OWL	71 (48%)	77 (52%)
Fall 2006	94	OWL	54 (57%)	40 (43%)
Fall 2007	71	MasteringChemistry	43 (61%)	28 (39%)
Spring 2008	146	CATALYST	79 (54%)	67 (46%)
Fall 2008	68	SmartWork	27 (40%)	41 (60%)

In preparation for a chi-square (or Pearson Chi-Square) test, the following correlation table was developed. Over all eight semesters 408 of 738 students were successful (55.3%), therefore the second number shown in each box is the expected number of students to be successful and unsuccessful based on the individual class size.

Table 9: Chi-Square Correlation Table

Class	Successful	Unsuccessful	Class Total
Fall 2004	30 27.65	20 22.35	50
Spring 2005	79 63.04	35 50.96	114
Fall 2005	25 25.99	22 21.01	47
Spring 2006	71 81.84	77 66.16	148
Fall 2006	54 51.98	40 42.02	94
Fall 2007	43 39.26	28 31.74	71
Spring 2008	79 80.74	67 65.26	146
Fall 2008	27 37.60	41 30.40	68
Total of All Classes	408	330	738

Analysis with a chi-square test yields the results that several of these classes vary from the expected amounts significantly.

Table 10: Results of the Chi-Square Test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	20.525[a]	7	.005
Likelihood Ratio	20.797	7	.004
Linear-by-Linear Association	6.940	1	.008
N of Valid Cases	738		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 21.02.

Table 11: Additional Symmetric Measures of the Chi-Square Test

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	**Spearman Correlation**	.095	.037	2.585	.010[c]
Interval by Interval	**Pearson's R**	.097	.036	2.645	.008[c]
N of Valid Cases		738			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.
c. Based on normal approximation.

Microsoft's Excel® program performs a chi-square analysis, but reports a percentage that estimates the likelihood the deviation from the expected values is attributable to chance (the higher the percentage, the more likely the deviation from the expected values is the result of various random factors). This offers a different way of viewing the data, and I choose to compare each class individually to the expected number of students for that particular class. The expected values are again the second number shown in each box.

Table 12: Results of Chi-Square Test as Chance by Class

Class	Successful	Unsuccessful	Chance Percent
Fall 2004	30 27.65	20 22.35	50.4 %
Spring 2005	79 63.04	35 50.96	0.2 %
Fall 2005	25 25.99	22 21.01	77.1 %
Spring 2006	71 81.84	77 66.16	7.3 %
Fall 2006	54 51.98	40 42.02	67.5 %
Fall 2007	43 39.26	28 31.74	37.2 %
Spring 2008	79 80.74	67 65.26	77.2 %
Fall 2008	27 37.60	41 30.40	0.9 %

Viewing the results in this manner makes it easy to see that the unusually high number of successful students in the Spring 2005 class is not at all likely to be the result of random chance, as well as the unusually low amount of successful students in the Fall 2008 class. Furthermore, placing the actual counts into a bar graph makes the extent of the departure from the normal trends even more clear as the Spring 2005 class has more than double the successful students as unsuccessful students, and the Fall 2008 class has noticeably more unsuccessful students than successful students.

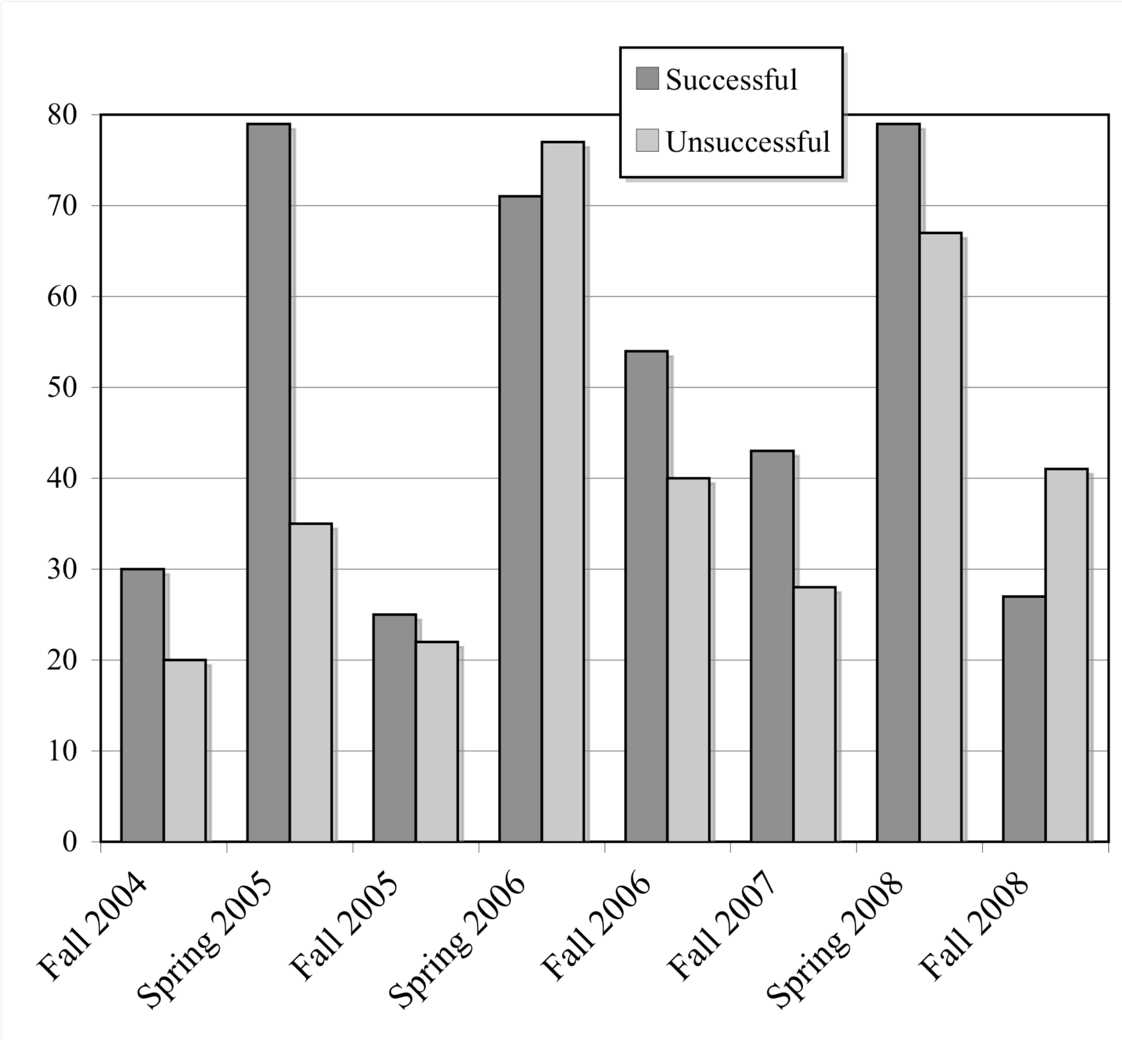

Figure 1: Successful/Unsuccessful students by semester.

CHAPTER 5

CONCLUSIONS

My null hypotheses at the onset of this study were:

1. There is no statistically significant difference between the classes that did not use an online homework system and those that did.
2. There is no statistically significant difference between any of the classes that used any of the four online homework systems.

At this time I feel there is reason to consider rejecting both hypotheses, but I would not claim the results of my groundbreaking study to be cause enough to completely reject these hypotheses.

Related to my first hypothesis, the ANOVA returns a statistically significant lower semester mean grade for the Spring 2005 class than for the two highest fall classes. Although this is a significant difference, it is showing the Spring 2005 class is the lower bound of all the class means. As it is not statistically significantly lower than the other five classes I hesitate to reject the first hypothesis based on this result alone.

Compounding this case, the Spring 2005 class has a significantly higher amount of successful students, inferring that this class saw an atypically large amount of students making lower successful grades: C's. As this class started with a statistically typical prior chemistry understanding (as per insignificant CA Dx ANOVA results), I am left to question if this was such an atypical class that it might be an outlier data point, and thus the sole objection to my first hypothesis might be excusable. Additionally, it does not match the trends already seen in the literature for similar studies, so it may be valid to conclude the Spring 2005 class results are excusable, and my first hypothesis should not be rejected. Further studies comparing paper and pencil homework systems to Web-based homework systems within chemistry classes will be needed to corroborate or refute the possibility of the Spring 2005 data being excusable outlier data.

Related to my second hypothesis, there is no statistical test result that demonstrates any of the Web-based systems statistically affect students' semester grades differently from another, but I think there might be a trend in the data when it come to the results of the Fall 2008 class.

The CA Dx results would indicate that the students in the Fall 2008 class started with an unusually high prior chemistry understanding, but by the end of the semester did not retain the position of the highest class semester mean, and in fact only had a mean higher than two other spring classes.

Adding to this alarming trend is the fact that many more students were unsuccessful than successful in the Fall 2008 class; a ratio not seen in any of the other classes. Overall I feel I would suggest any chemistry instructor to carefully consider the options before choosing SmartWork as the Web-based homework system for a class.

I do fully acknowledge when studying complex situations like a classroom, and studying even more complex subjects like students there is no likelihood that this perceived trend with the Spring 2008 class is resultant entirely because of the homework system. Still it is enough to cause me to consider my second hypothesis may not be sound. Further studies including SmartWork would certainly be needed to establish if I am warranted in considering my second hypothesis should be rejected.

Unfortunately, this study was proposed after the data for most of the classes were collected, so there is no way to add in an additional diagnostic exam, or to poll the students psychologically to determine how likely it is that the differences seen are the direct result of the homework system.

Subsequent studies on this topic would want to plan more than one diagnostic exam at the beginning of each class (one to measure prior chemistry understanding and one to measure thinking aptitude - like the GALT exam), and it would be better to keep one homework system for both one fall and one spring class. This would promote the additional comparison of fall-to-fall and spring-to-spring classes without having to consider the effect of fall to spring comparisons.

Furthermore, a self-report survey of the students to determine if they are taking the course for the first time or repeating the course and who their previous instructor was might allow for repeating students to be factored out of the analyses, in case the scores of these repeaters would effect other results.

Most instructors know that not every successful student is successful on the homework and not every unsuccessful student does unsuccessfully on the homework, so it might be enlightening to determine if there is a correlation specifically between a student's success on the homework and the student's success in the class. For this study homework grades were a part of the final semester grades, so it is probable that there would be no unique trend discovered, but it might be worth performing such an analysis just to determine if the results are similar or not if the study can be designed to specifically gather this information..

Moreover, it would be best if the classes could maintain identical course sequences, but all educators know this is very difficult to control, especially as, for example, universities often have a spring break in the middle of the spring semester but Thanksgiving break is close to the end of the fall semester.

I do invite anyone that has the opportunity to design a study such as mine to continue the work so that the chemistry field can benefit from knowing if Web-based homework systems are not only a valid and helpful tool, but reliable as well, regardless of the origin and features of the homework system.

REFERENCES

1 Sanders, R. (2005). Redesigning introductory Spanish: Increased enrollment, online management, cost reduction, and effects on student learning. *Foreign Language Annals, 38(4)*, 523.

2 Cole, R., & Todd, J. (2003). Effects of Web-based multimedia homework with immediate rich feedback on student learning in general chemistry. *Journal of Chemical Education, 80(11)*, 1338-1343.

3 Penn, J., Nedeff, V., & Gozdzik, G. (2000). Organic chemistry and the Internet: a Web-based approach to homework and testing using the WE_LEARN system. *Journal of Chemical Education, 77(2)*, 227-231.

4 Allain, R., & Williams, T. (2006). The effectiveness of online homework in an introductory science class. *Journal of College Science Teaching, 35(6*), 28-30.

5 Zerr, R. (2007). A quantitative and qualitative analysis of the effectiveness of online homework in first semester calculus. *Journal of Computers in Mathematics and Science Teaching, 26(1)*, 55-73.

6 Hauk, S., & Segalla, A. (2005). Student perceptions of the Web-based homework program WeBWorK in moderate enrollment college algebra classes. *Journal of Computers in Mathematics and Science Teaching, 24(3)*, 229-253.

7 Russell, A. (1994). A rationally designed general chemistry diagnostic test. *Journal of Chemical Education, 71(4)*, 314-317.