

Organic Chemistry and the Internet: A Web-Based Approach to Homework and Testing Using the WE_LEARN System

WE_LEARN system is now owned by John Wiley & Sons and has been rebranded "CATALYST" and is available within WileyPLUS.

John H. Penn* and Vincent M. Nedeff

Department of Chemistry, West Virginia University, Morgantown, WV 26506-6045; *JPenn2@wvu.edu

Gloria Gozdzik

Horizon Research Consultants, Inc., Morgantown, WV 26505

The computer and the Internet have proven to be extremely versatile tools for a variety of applications. A complete citation listing of the uses of computers in chemical education is beyond the scope of this manuscript. However, a sampling of the recent uses of computers in chemical education reported in this *Journal* is given here to indicate some of the techniques now available (e.g., application of hypermedia and other Internet methods [1–3], kinetics [4, 5], X-ray powder diffraction simulation [6] mathematical software for chemical applications [7], visualization of structures [8, 9], course delivery over the Internet [10], course delivery using Internet techniques [11]).

Although computer methods have achieved a measure of success when viewed from an instructional standpoint [12], we desired a computer-assisted tool that would enable the instructor to become part of the educational process throughout the entire course. A necessary component of any system for the instructor is a flexible tool that can be modified to meet the individual needs of both students and instructors.

Several design features are critical to the success of the system. The learning system must provide immediate and adequate feedback. For the student, immediate feedback is well known to be essential for enhanced learning (13–17). For the instructor, feedback would allow for changes in instructional methodology or re-explanation of critical concepts. Normally, feedback to the instructor is available only after the completion of an in-class examination. As a second critical design feature, the system must engage the students in an active learning mode, since it is well known that an active learning environment is superior to a passive one (18). A third design goal is to ensure that learning occurs in both similar and dissimilar environments (19, 20). In the similar-learning environment, all problems related to a certain topic, area, or idea are grouped together. Students are able to easily identify the concepts being tested or assigned and are able to improve their understanding of or proficiency in these concepts. An enhanced learning situation would be created by having a practice situation whereby students could personally assess their learning before an exam. This corresponds to the dissimilar learning situation, as various concepts, topics, and problems are randomized out of the rote-memorization context. A final design feature is to ensure a high degree of flexibility within the system, since no two students learn the same way and no two instructors teach the same way.

These requirements are easily met with today's technology by combining Internet capabilities with the widespread availability of computers to students. We have developed the Web-

based Enhanced Learning Evaluation And Resource Network (WE_LEARN) system following these design requirements. This system comprises a software engine (designed as an Internet server) and a database of questions that is easily controlled and manipulated by the software engine. Almost instantaneous feedback is provided to students when they "submit" answers to a series of questions provided via an Internet server. Feedback to the instructor is provided by information concerning student performance on questions related to class material. The system provides an active learning mode, since students select the correct answer(s) from a list of potential answers. Many individualized "modules", which focus on similar concepts, create a similar learning environment. The software maintains the capacity for randomly selecting questions from a larger subset of questions already learned within the individualized "modules" to provide a dissimilar learning environment. Flexibility for students is achieved by using the World Wide Web for distribution of materials. Web principles allow anyone to use the WE_LEARN system at any place (e.g., on campus, in the privacy of one's own room) at any time. Research has shown that students differ from university administrators and professors in the times and places used for study (21). We describe here our development and implementation of the first prototype of the WE_LEARN system for a section of organic chemistry.

Experimental Methodology

Software Engine and Internet Server

The Cyberexam program from Virtual Learning Technologies, LLC (1401 20th Street South, Suite 300, Birmingham, AL 35205; <http://www.vlearning.com>) was used. This engine provides an Internet server and maintains a database of statistics on students and their assignments. This server was chosen because it allows for any HTML commands (e.g., graphics files, multimedia support), online grading, a wide range of question types (multiple-choice, multiple-select, true/false, short-answer, fill-in-the-blank, selection from a list), and random test generation. The software also provides flexibility for a large number of test administration techniques.

Question Database

A multiple-select question database was created, similar in style to the homework assigned from the textbook that was then used in our classroom. In fact, the vast majority of the questions were taken directly from the textbook to ensure

that students who did not elect to utilize the WE_LEARN system did not suffer an unfair disadvantage relative to students who did utilize the new technology. Multiple-select questions were chosen as the primary question-delivery mode because organic chemistry requires a large amount of structural information that is best represented in a graphical format. Although the system had the capability for the students to draw structures and submit them via Internet methods, chemical structure drawing and submission was not the aim of our study. Therefore, structures were generated for the students and presented as part of a multiple-select format. This format (where zero, any, or all of the answer choices are potentially correct) was chosen to prevent students from eliminating answers simply by choosing one correct answer (i.e., all answers must be considered). Incorrect (distracting) answers were generated by choosing the incorrect answers that students had submitted in written homework assignments during earlier semesters of this course.

Assignment Methodology

The homework was assigned in both a similar learning environment and a dissimilar learning environment. In the similar learning environment, examples of an assignment might be a series of questions relating to the nomenclature of aromatic compounds or to the Grignard reactions of aldehydes and ketones. These assignments were made after the material had been covered in the course lecture. Both the question order and the presentation order of the answers were randomized to lessen the possibility that answers were memorized. Each assignment was structured to enable students to complete the assignment in approximately 10 minutes. This would allow students without their own Internet access to complete assignments in university computer labs, where there is a high background noise and activity level. When the assignments were completed, students submitted them for grading by clicking on the "Submit" button. They received immediate feedback about their correct and their incorrect responses. Students were permitted to do the assignment a maximum of 10 times. Each time the assignment was undertaken, a different set of questions in a randomly generated order was obtained. For each in-class exam, there would be a minimum of 8 individualized assignments that contained an average of 120 questions.

Before each exam, students were moved into the dissimilar learning environment. This learning environment was created by placing all the questions previously assigned as single-concept questions (e.g., nomenclature, Grignard) into a "database" of questions consisting of all the concepts taught for that exam. When students accessed this practice test, the Cyberexam server would select 10 questions at random from the practice test "database". These practice tests could cover multiple concepts and material unrelated to another question selected for students' assessment. Knowledge of and proficiency in the material was assessed before the actual examination by submitting responses to these questions for grading. The computer immediately graded the answers and indicated which questions were answered correctly and which incorrectly. Because the material included in each practice exam was random, we hypothesized that if students consistently received a grade of 100% on each practice test, then they would perform well on the in-class examination.

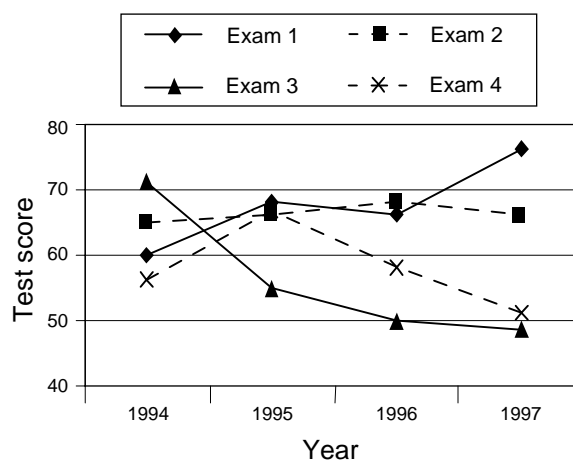


Figure 1. Class averages for the fall semesters, before implementation of the WE_LEARN system. The exam format changed from fill-in-the-blank to multiple choice in 1997.

Samples of the question methodology and a sample practice test are maintained at our WE_LEARN server (<http://www.access3.wvu.edu>). Access to these sample assignments is available by using the account loginID of "guest" and password "guest". The server is case-sensitive. Therefore, be sure to use lowercase letters.

Results and Discussion

Higher Test Scores

A critical issue in this discussion is the change in exam style from a fill-in-the-blank testing strategy, which was used in previous years, to the multiple-choice or multiple-select testing strategy used in the WE_LEARN system. Despite a large volume of literature to the contrary (22), many students and instructors believe that multiple-choice testing is easier and less educational than a fill-in-the-blank examination. For this reason, the testing strategy for this course was changed one semester early to verify that similar grades are obtained in multiple-choice tests and fill-in-the-blank tests.

Figure 1 depicts data on the scores obtained in multiple-choice and fill-in-the-blank exams. Because the same textbook had been used for a number of years in this course and the course was structured similarly each year, it was possible to compare the tests given during a period of several years. In the fall 1997 semester, a multiple-choice testing format was used. As seen in Figure 1, there was no significant change in the class averages for each test in the fall 1997 semester when compared to the results of previous years. Similar trends were observed. Only exam 1 showed an increase. However, in 1997 this exam covered only the first two chapters of the textbook, as opposed to three chapters in the preceding years. For all other exams, the class average in the multiple-choice exam format was less than or equal to the class average in previous years. Thus, these results are in agreement with previous studies indicating that properly designed multiple-choice tests are as effective as fill-in-the-blank methods for evaluating student knowledge (22).

Examining the class average for the entire semester provides a further check of these data. The semester averages in the fall semesters (prior to WE_LEARN) are 63, 64, 60, and

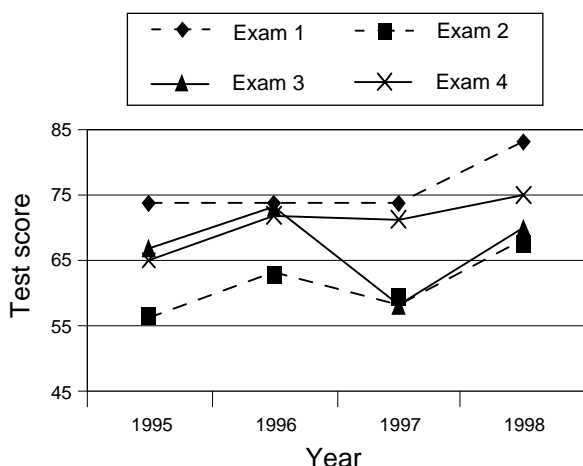


Figure 2. Class averages for spring semesters. The WE_LEARN system was implemented in spring 1998.

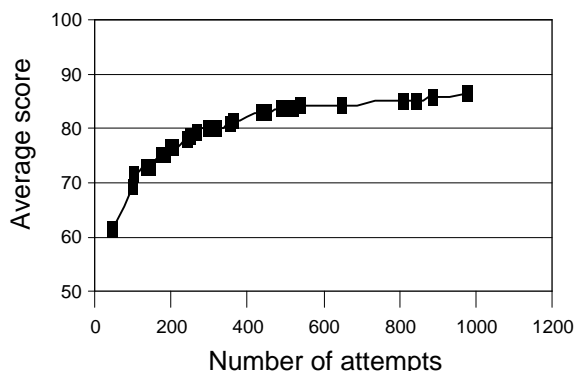


Figure 3. Average score on practice test 4 as a function of the number of attempts.

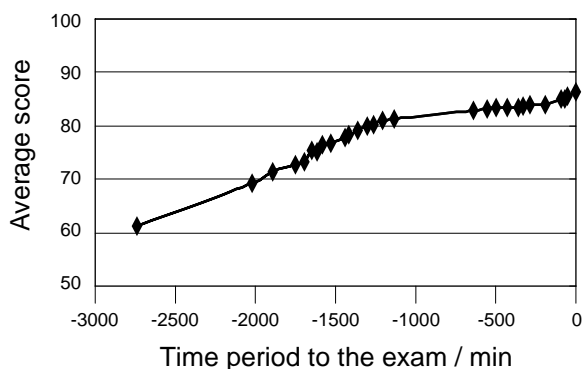


Figure 4. Average score on practice test 4 as a function of time remaining until the exam (time 0).

60 for the years shown in Figure 1. This is in agreement with our assertion that no increase in class average occurred in shifting from a fill-in-the-blank testing format to the multiple-choice testing format.

Data for the spring '98 semester, in which the WE_LEARN system was implemented, are shown in Figure 2. For all exams, an increase in the class average of 5–10% is observed. This occurred despite the multiple-select format of the exams of the WE_LEARN system. The variability in scores for exam

3 (1996) is related to subtle changes in exam coverage in 1996 and 1997 and the coverage of easier material in the 1996 spring semester. Examining the class average for the entire semester provides a further check of these data. In contrast to the data for the fall semesters, the semester averages in the spring semesters (Fig. 2) are 66, 70, 65, and 74 in 1995–1998, respectively. The increased average of 74 for the spring '98 semester clearly shows that the change to the WE_LEARN system had a positive impact on the class.

The apparent trend toward higher test scores could be challenged by questioning the makeup of the in-class exams and the source of the questions for these exams. Prior to the introduction of WE_LEARN, about 75% of questions on the in-class exam came directly from the assigned homework questions. Answers for these questions were readily available at the back of the textbook and in a solutions manual that could be purchased by the students and was also on reserve in the library. With the introduction of WE_LEARN, the percentage of questions taken directly from homework assignments was reduced to 62.5% after the first exam. This percentage has been further reduced with no concomitant decrease in exam scores. We believe that the increase in test scores originates from an enhanced conceptualization of key concepts. However, we cannot rule out the explanation that we have enabled students to study better because the answers to all the assigned problems are gathered together for more ready access and subsequent study. In either case, the system was responsible for the trend toward higher test scores.

Average Practice Exam Score Increases Prior to an In-Class Exam

Feedback on student performance is readily obtained from this system and provides a mechanism to evaluate system effectiveness. In addition to our average test score data, a cumulative average for all students who took a given assignment is available. As an alternative method for evaluating system effectiveness, we recorded the average score for a practice test as a function of time and of the number of completed attempts on the practice exam. Representative data are shown in Figures 3 and 4 for exam 4. Although these data are for only one exam, similar trends were observed for all other in-class exams throughout the semester.

A plot of the cumulative average on the practice test for exam 4 as a function of the number of times the practice test was taken is shown in Figure 3. The average score starts at 61% and reaches an asymptotic value of ca. 90%. This indicates that performance on the practice exam improves with the number of attempts students make on the system.

An alternative but similar view of the data is obtained by comparing the average practice exam score to the time prior to the in-class exam (Fig. 4). The same trend is observed. The practice exam average score rises from an initial value of 61% to a value approaching 90% as the in-class exam time approaches. Although only one practice exam's data are presented, similar data were obtained for all other practice exams administered during the semester.

One might ask if these data are exceptional, since an increased amount of review of the material *should* result in an increase in the exam scores. The computer system provides a means for documenting the amount of review time, which was not possible under the old system. We believe that the

ability to document student effort affords a pedagogical advantage over previous methods. The cumulative class average and the cumulative amount of time spent on review could provide an important diagnostic for the instructor. In the same way that cumulative class data are available, the Web-based method described here also allows for examination of the effort of individual students.

Growth in Student Usage as a Function of Time

Data related to the usage of the practice exam feature for this course are shown in Table 1. These data are presented to show how the students themselves reacted to the system. The increase in level of usage was not an effect mandated by the instructor, nor was it induced by incentives (e.g., extra credit). The only incentive offered to use the system came at the beginning of the semester. Students could earn up to a maximum score of 1/4 of a letter grade, based upon their maximum percentage score for each assignment. There was no subsequent modification to this incentive, nor were there announcements about it during classroom lectures.

Each practice exam represents 10 questions selected at random from the total database of questions related to the material to be contained in the real exam. Each attempt represents a student completing a practice test and submitting it for grading. The trend is clearly toward an increasing number of attempts prior to an exam. Only practice exam 3 showed a decrease in the number of attempts. This decrease is related to an administrative change in the location of the Cyberexam server to a location that gave a lower overall rate of data transmission over the Internet. This resulted in long loading times (ca. 5–10 minutes per 10-question practice exam during peak usage periods). The overall system response became so slow that both instructor and students became frustrated and used it less often than was desirable from an instructional standpoint. This problem was rectified before the next exam (exam 4) by placing the server on a faster Ethernet hub.

The percentage of the class using the WE_LEARN system shows a trend similar to the trend in total number of attempts on the practice exam. The percentage of the class that was

Table 1. System Usage, Spring '98

Practice Exam No.	Attempts (No.)	Percentage of Users
1	578	55.5
2	891	57.0
3	458	49.2
4	1000	65.6
5	1144	66.9

using WE_LEARN increased as the semester proceeded (Table 1). The downturn in usage of the system for practice exam 3 must be related to the poor system response described above. Since there was no instructor-driven reason for the students to use this system (no mandate or incentive), we conclude that these usage statistics show that the students themselves found the system to be worthwhile and used it as a study aid.

Qualitative Data Relating to Faculty/Student Interactions

In addition to the positive changes in the quantitative assessment tools originating from the class after implementation of WE_LEARN, qualitative data also indicate positive changes in the learning environment. Qualitative indicators include the number of students coming to the instructor's office or to a help-session during the week of an in-class exam,

the types of questions asked, and the air of student confidence during the in-class exams.

Before implementation of WE_LEARN, long lines of students outside the instructor's office were typical during an exam week. Although no statistics on the number of students requesting help were kept, an estimate of the average number of students coming to the office for help is 20–30, or about 15% of the class. An estimate of the time required to answer the questions from these students is 20 hours. After the introduction of the WE_LEARN system, a dramatic change in the number of students requiring office help took place—so dramatic, that quantifying its magnitude was rather easy. After WE_LEARN was implemented, an average of only 3 students, requiring only 15 minutes of total time, came for office assistance before each exam. A similar striking effect was noted in the attendance at help sessions, which have been given on the night before an exam for many years. Prior to the introduction of WE_LEARN, typical attendance at the help session was ca. 75% of the class enrollment. After the introduction of WE_LEARN, attendance at help sessions dropped to ca. 33% of the class.

The types of questions asked by students were also of a different nature following the introduction of WE_LEARN. Before the implementation of the system, the majority of questions posed by students, both in the office and in the pre-exam help sessions, 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 now came from students with the strongest critical thinking skills and were very specific to the types of reaction conditions or subtle structural variations that might lead to a different set of products.

A final qualitative indication of the change in student attitude and perception was the change in classroom atmosphere during in-class exams. Prior to WE_LEARN there was an atmosphere of tension and anxiety. Almost all students utilized a minimum of one hour to complete the exam and ca. 10% of the class utilized the entire two-hours allocated for the exam. This distribution of time to complete the exam occurred in both the fill-in-the-blank questions of prior years and the multiple-choice questions of the Fall '97 semester. After the introduction of the WE_LEARN system, most students took much less than one hour to complete the exam and no one took the entire two hours. The faces of the students appeared much more confident. These data indicate that it is the WE_LEARN system that is responsible for the change in the tension level and time required to complete the exam and not the change to a multiple-select format.

Summary and Conclusions

Together, these quantitative and qualitative data clearly indicate that a positive change occurred in the classroom upon the introduction of the Web-based Enhanced Learning Evaluation And Resource Network (WE_LEARN) system. In comparison to previous years, students obtained higher test scores and were more confident when taking exams. Students used the system as a learning tool, as demonstrated quantitatively by the increase in the average score on the practice exams and increased usage of the system as the semester progressed. Qualitative data on the types and number of questions asked

by students indicate that the students have a better feeling about their level of knowledge before exams when using the WE_LEARN system. It is tempting to conclude that the higher test scores indicate that the students learn more when using this system. However, this conclusion requires a firm definition of the meaning of learning and interpretation of variables that are beyond the level of the current study. We conclude that a positive change has occurred in the classroom environment and are seeking to further elaborate the reasons behind the positive results.

Acknowledgments

Many people have been involved behind the scenes in providing the technical expertise to accomplish the implementation of the WE_LEARN system. The West Virginia University Office of Academic Computing (Donald McLaughlin, Sherrie Kelly, and Bradley Forbes) have been extensively involved in the discussions concerning system requirements and hardware maintenance. Virtual Learning Technologies, LLC (Gary Hulse, James Johnson, and Morgan Sapp) was an invaluable asset during those moments when software problems threatened to ruin this classroom experiment. Finally, the 130 students in the major author's section of Organic Chemistry (Spring '98) are gratefully thanked for their efforts and patience when problems arose.

Literature Cited

1. Tissue, B. M. *J. Chem. Educ.* **1996**, *73*, 65–68.
2. Tissue, B. M. *Trends Anal. Chem.* **1997**, *16*, 490–495.
3. Smith, S.; Stovall, I. *J. Chem. Educ.* **1996**, *73*, 911.
4. Guzmán, M.; Ortega, P.; Vera, L. *J. Chem. Educ.* **1996**, *73*, 921.
5. Bruist, M. F. *J. Chem. Educ.* **1998**, *75*, 372–373.
6. Masson, B. L. *J. Chem. Educ.* **1996**, *73*, 918.
7. Smith, W. R.; Missen, R. W. *J. Chem. Educ.* **1997**, *74*, 1369.
8. Crouch, R. D.; Holden, M. S.; Samet, C. *J. Chem. Educ.* **1996**, *73*, 916.
9. Childs, S. L.; Hagen, K. S. *J. Chem. Educ.* **1996**, *73*, 917.
10. Liu, D.; Walter, L. J.; Brooks, D. W. *J. Chem. Educ.* **1998**, *75*, 123–124.
11. Matthews, F. J. *J. Chem. Educ.* **1997**, *74*, 1011–1014.
12. Mehan, H. *Anthropol. Educ. Q.* **1989**, *20*, 4–22.
13. Dubinskas, F. A. *NAPA Bull.* **1993**, *12*, 1–11.
14. Clark, K.; Dwyer, F. *Int. J. Instructional Media* **1998**, *25*(1), 55–63.
15. Carter, J. *Comput. Teach.* **1984**, *12*, 53–55.
16. Alessi, S. M. *Computer-Based Instruction Methods and Development*; Prentice Hall: Englewood Cliffs, NJ, 1991.
17. Mory, E. *Educ. Technol. Res. Dev.* **1992**, *40*(3), 5.
18. Lagowski, J. J. *J. Chem. Educ.* **1998**, *75*, 425–436.
19. Osin, L.; Lesgold, A. *Rev. Educ. Res.* **1996**, *66*, 621–656.
20. Farnham-Diggory, S. *Rev. Educ. Res.* **1994**, *64*, 463–477.
21. Anderson, K.; McClard, A. *Anthropol. Educ. Q.* **1993**, *24*, 159–177.
22. Haladyna, T. M. *Developing and Validating Multiple-Choice Test Items*; Lawrence Erlbaum Associates: Hillsdale, NJ, 1994.