ONLINE LEARNING AND TIME-ON-TASK: IMPACT OF PROCTORED VS. UN-PROCTORED TESTING

Gregory S. Wellman
Assistant Professor of Pharmacy
College of Pharmacy
Ferris State University
Email: wellmang@ferris.edu

Henryk Marcinkiewicz
Associate Vice-President for Academic Affairs
Pennsylvania College of Technology
Email: hmarcink@suscom.net

ABSTRACT
The purpose of this study was to evaluate the impact of proctored versus un-proctored testing on practice time and learning for a self-study, online course in medical terminology. Participants included 120 college students in a pharmacy curriculum. Subjects were randomized to one of three groups utilizing textbook with proctored assessment, online instruction with proctored assessment, or online instruction with un-proctored assessment. A pre-test and post-test was administered to all three groups to evaluate learning. Content page “hits” and practice quiz access was tracked for participants. Learning, as defined by pre/post-test change score, was greatest in the presence of proctored assessment (online and text) (p = 0.027). In addition, use of practice quizzes had a stronger relative correlation with learning (r = 0.401; p<0.001) when compared to content page “hits” (r = 0.257; p = 0.024). Online course content paired with meaningful time-on-task (e.g. practice quizzes) was most effective when paired with proctored assessment.

KEYWORDS
Distance learning, Computer-mediated instruction, Proctored testing, Un-proctored testing

I. INTRODUCTION
As educators adopt online instructional techniques, one of the challenges they face is assessing learner mastery of online course content. Online course management systems (CMS) can assist by facilitating the three important components of instruction: presentation of information, practice (a proof to oneself), and testing (a proof to a criterion). Course management systems provide a variety of tools to facilitate student practice through the use of discussion groups, chat, self-tests or quizzes. Online media are effective for content presentation, but the conditions for effective online practice and testing are still being determined. Two keys to effective practice are time-on-task and feedback, which help students master course content and prepare them to demonstrate their competency through testing.
Practice and testing are similar types of activities that differ principally in the value of the stakes involved. Practice activities have relatively low stakes for the student. This is in contrast to testing, which is typically a high-stakes activity that is tied directly to grades. The need for practice activities and feedback has a basis in the conditions suggested by Gilbert’s model of human competence [1]. If we consider learning the behavior of interest, the conditions necessary for competence are as follows:

1. **Information** (Data): The teacher communicates expectations for participation in instruction that match with the career goals and objectives of the students, as well as ongoing, relevant and frequent feedback to assist the student in understanding their adequacy of performance.
2. **Instrumentation** (Instruments): The teacher provides instructional media that match the learning style and capabilities of the student, including meaningful time-on-task.
3. **Motivation** (Incentives): The teacher provides incentives that match the motivational needs of the student.

Interestingly, we can see these conditions interwoven in the seven principles for good practice in undergraduate education articulated by Chickering and Gamson [2]:

1. Encourages contacts between students and faculty;
2. Develops reciprocity and cooperation among students;
3. Uses active learning techniques;
4. Gives prompt feedback;
5. Emphasizes time on task;
6. Communicates high expectations; and
7. Respects diverse talents and ways of learning.

When applying Gilbert’s model to online course design, it becomes important to include time-on-task, as well as relevant corrective feedback, to complement content delivery. For motivated practice to occur, there should be a link between time-on-task and success in the high-stakes activity of testing. This is supported by research showing time-on-task as a predictor of performance [3–9].

Traditional in-class testing provides face-to-face contact between student and instructor to allow for clarification of test instructions, direct supervision of students, and feedback on performance. Often, feedback and its value to learning are delayed because of the time required to grade tests and distribute results. Course management systems enable the instructor to test students asynchronously, and also to give immediate corrective feedback with online grade results. This can be done with a number of different testing formats, including multiple choice, true-false, and short answer. Asynchronous quizzing or practice, paired with automated grading, can give the student immediate feedback on performance while the questions and content are still fresh in their mind, thereby enhancing learning. There is, however, a paucity of research examining the impact of proctored versus un-proctored testing on time-on-task and learning in an online teaching environment.

**II. BACKGROUND**

As part of the first year of the pharmacy curriculum, students are required to complete a self-study of medical terminology, which is administered through a Practice Skills Lab. As part of a curricular revision of the lab, an online self-study content module with asynchronous delivery and online testing in medical terminology was created. It was felt that an online, CMS-based instruction module on medical terminology would provide:
(a) Flexibility of access for students, since they had access to web-enabled computers in the College of Pharmacy and university library, in addition to those who had home computers.
(b) Meaningful time-on-task using self-tests and practice quizzes to complement the study of material.
(c) Efficiency in managing the feedback and scoring of practice quizzes for a large number of students.
(d) An opportunity to allow students to conduct their learning in a completely asynchronous format.

The overarching questions that emerged were to what extent would students benefit from the use of online self-tests and practice quizzes and would un-proctored, online quizzes be as effective as the in-class ones?

The objectives of this study were to:
(a) Compare the effectiveness of self-directed textbook based learning to self-directed online learning;
(b) Identify the relationship between time-on-task, as defined by the use of practice quizzes, and performance on objective measures (testing); and
(c) Assess the impact of proctored versus un-proctored testing on practice and learning.

III. METHOD

A. Participants
Subjects included 120 college students in a Doctor of Pharmacy program. Of these, 83 were female and 37 were male. Students were enrolled in a Practice Skills Lab in the first of a four-year program. Students in this program are required to take a standardized aptitude test (Pharmacy College Admission Test, The Psychological Corporation, San Antonio, TX), as well as complete at least two years of prior academic study at the college or university level.

B. Materials
An instructional module in medical terminology was developed on WebCT. The module began with rules for word construction and the use of prefixes and suffixes. Chapters were then organized according to the human anatomical organ systems (e.g. gastrointestinal, respiratory, cardiovascular, integumentary, eyes, etc.). Each chapter featured key root words, examples of prefix-root-suffix combinations, and illustrations and examples of how the words are used in everyday medical practice. Practice quizzes, based on multiple-choice questions, were designed for each chapter to give the student the opportunity to test their mastery of the medical terms. The practice quizzes consisted of 5 to 6 questions that were designed to provide samplings of the complete database of medical terminology questions compiled for each chapter. Corrective feedback was designed into the practice quiz questions, and students were able to retake practice quizzes an indefinite number of times.

In addition to the online instruction, a medical terminology book was selected for use in the study that was organized in a similar fashion. The textbook contained robust examples and illustrations. At the end of each chapter, exercises were available (with answers in the appendix) for the student to test their mastery of terms. Exercises were fill-in, multiple-choice, and matching.
C. Pre-Test
Students completed an unannounced pre-test that was administered in the first week of the semester. The pre-test was proctored and comprised of 40 multiple-choice questions derived from medical terms that would be covered by the book or online module. Students were informed of the nature and design of the entire project and given unlimited time to complete the pre-test. They were informed before taking the pre-test that they would receive an automatic 25 out of 25 score (equivalent to ¼ of an exam grade for the course) for taking the pre-test in a conscientious fashion, without regard for their actual performance on the pre-test. It was further explained that any indication that they were simply answering in a pattern (e.g. a, b, c, d, c, b a) would result in forfeiture of these points.

D. Procedure
Students were randomly assigned to one of three groups that varied based on the format for content delivery and the presence or absence of proctoring.

   a. **Online self-study with online proctored quizzes (40 students):** Students were instructed in the use of the online CMS and given access to the instructional module on medical terminology. The module was accessible, with student-specific login via any web-enabled computer terminal from campus or home. Written instructions were given about the timing of chapters to be studied during the semester. Students were also instructed on the availability and use of chapter practice quizzes that could be taken an unlimited number of times. This group was given two in-class, online, proctored quizzes covering the assigned chapters that counted toward their grade in the course. Students were assigned a specific time in the computer lab to complete their quizzes. College faculty verified student identity and supervised their login and completion of each of the two in-class quizzes.

   b. **Online self-study with un-proctored quizzes (40 students):** Procedures were identical to the first group except that students in this group were given two online, un-proctored class quizzes, covering the assigned chapters that counted toward their grade in the course. Students were given a 24-hour block of time to complete their quizzes, which began immediately after the other two groups completed their quizzes. Students were instructed that they were on the “honor” system to take the quiz on their own, and that they could do so at any web-enabled computer from campus or home using their course login.

   c. **Textbook-based self-study from with proctored quizzes (40 students):** Students in this group were provided with a self-study medical terminology text and instructed on its use for self-directed learning of the material. Textbooks were provided to students without charge. Written instructions were given about the timing of chapters to be studied during the semester. Students were given a list of specific exercises to be completed for each chapter, as well as practice quizzes (paper and pencil) written by the course instructor. This group was blocked from accessing the online CMS medical terminology module. The group was given two in-class, proctored quizzes covering the assigned chapters that counted toward their grade in the course. The quizzes were administered in a paper-and-pencil format.

Proctored and un-proctored class quizzes (2 per group) for the CMS groups were comprised of 30 multiple-choice questions, derived from the question databases used for the practice quizzes. The time allotted to take each quiz was limited to 20 minutes. Students were informed of the nature of the proctored and un-proctored class quizzes, and that they counted toward their grade in the class. The first of the two quizzes, covering approximately half of the medical terminology course content, was given in the 6th week of the 16-week semester. The second quiz, covering the second half of the material, was given in the 13th week of the 16-week semester.
E. Post-Test
The 40 multiple-choice question post-test was administered in the 16th week (during the assigned final exam time), allowing a three-week time-lapse between the last quiz and the post-test. Students were informed in the first week that they would be taking a post-test, which they were reminded of one-week prior to the post-test. They were told not to study for the post-test, and that they would again be given an automatic 25 out of 25 for conscientiously completing it, without regard for their actual performance on it. Administration of the post-test during “finals” week, along with instructions that their score was “automatically” 100%, was meant to provide sufficient distraction (preparing for other course final exams), and grade assurance (automatic score) to prevent students from studying for the post-test.

The proctored pre and post-tests were scored by hand, to give the course instructor the opportunity to observe any patterned answering. In addition, a few very common medical terms were included, again to signal a student who might be indiscriminately marking answers on the pre and post-test. Raw scores (out of 40) were converted to percentages.

F. Dependent Measures
Learning was defined by change score for the pre/post-test (% post-test minus % pre-test). Time-on-task was defined, for the CMS groups, as web page “hits” on content and quiz pages, and number of attempts at practice quizzes contained within each chapter. Meaningful time-on-task data could not be collected for the book group. Proctored and un-proctored, as well as practice quiz questions, were derived from the same database to maintain a strong association between time-on-task (doing exercises and practice quizzes) and the potential for a higher individual student grade for this part of the course. This was done to encourage students to continue to utilize the study mode (book versus CMS) provided and to observe the relationship between time-on-task and learning, as defined above.

G. Data Analysis
An individual group sample size of 40 was estimated using a multiple comparison’s calculation with a priori alpha = 0.05; effect size = 0.20; and power = 0.80. Analysis was done using SPSS for Windows, version 10 (SPSS, Inc., Chicago, Illinois). The a priori level of significance used was 0.05.

IV. RESULTS
Of the 120 students enrolled and pre-tested at the beginning of the semester, and randomized to the three groups of 40, 118 completed the course medical terminology module and post-test (two students withdrew from the program). One student’s post-test appeared to have confusing and indiscriminate answers, and was the only test that included incorrect answers on the “common” medical terms. Scores for this student and the two students who withdrew from the program were excluded from analysis. The final distribution of usable data was as follows: Textbook group with proctored quizzes — 40 responses; online group with proctored quizzes — 38 responses; and online group with un-proctored quizzes — 39 responses (total n = 117). Of the total, females made up 69% (81) of the subjects and males made up 31% (36).

A. Prior Knowledge
As mentioned, pharmacy students are required to complete a minimum of two years of “pre-pharmacy”, college level coursework in the areas of chemistry, biology, English composition and the humanities. In
addition, students must complete the Pharmacy College Admission Test (PCAT), which measures verbal ability, quantitative ability, reading comprehension, and knowledge in biology and chemistry. Groups did not differ on cumulative pre-pharmacy grade point average (p = 0.989) or cumulative PCAT score (p = 0.786) as tested by analysis of variance.

B. In-Class Quizzes
As expected, deriving quiz questions directly from online and off-line (textbook group) practice quizzes and exercises (book group), yielded high mean averages for all three groups for the two proctored or unproctored quizzes that were administered as part of the class during the semester (online with unproctored quiz group — 96.3%; online with proctored quiz group — 93.3%; and book with proctored quiz group — 91.4%). An analysis of variance was conducted to determine if mean quiz scores (for the two class quizzes) differed between the three groups. The null hypothesis was rejected (F = 4.855; p = 0.009), supporting the observation that there were differences in mean class quiz scores between the three groups. Post hoc analysis, using Tukey HSD, revealed two subset groups that included:

1. Subset One: Online with un-proctored quizzes and online with proctored quizzes, and
2. Subset Two: Online with proctored quizzes and book.

As a result, students in the online-with-un-proctored-quizzes group performed better in unsupervised class quizzes than students in the book group did in proctored-class quizzes. There was no statistical difference in class quiz performance when comparing students in the online-proctored group to the online-unproctored or book, proctored groups.

C. Change Score
As discussed, all students also took a proctored pre (week 1) and a post (week 16) test to assess learning. Mean change score (post-test % minus pre-test %) for each group is illustrated in Table 1 and Figure 1. The greatest improvement in change score was seen in the online-with-proctored-quiz group, followed by the book-with-proctored-quiz group, then the online-with-un-proctored-quiz group. An analysis of variance was conducted to determine if mean change score differed between the three groups. The null hypotheses was rejected (F = 3.746; p = 0.027), supporting the observation that there were differences in mean change score between the three groups. Post hoc analysis, using Tukey HSD, revealed two subset groups that included:

1. Subset One: Online with proctored quizzes and book; and
2. Subset Two: Online with un-proctored quizzes and book.

As a result, students in the online-with-proctored-quizzes group had a greater improvement in change score when compared to the online-with un-proctored-quiz group. The improvement in change score for students in the book group could not be distinguished, statistically, from either online group. Female students (mean change score = 29%), across all three groups combined, had greater improvements in change scores than males (mean change score = 23%). This difference was significant using the t-test for independent samples (p = 0.017).
Table 1: Pre-Test, Post-Test and Change Score Averages

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre-Test Score</th>
<th>Post-Test Score</th>
<th>Change Score</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book: Proctored Quiz Group</td>
<td>40</td>
<td>50.8</td>
<td>76.6</td>
<td>25.8</td>
<td>14.1</td>
</tr>
<tr>
<td>CMS: Proctored Quiz Group</td>
<td>38</td>
<td>47.4</td>
<td>79.2</td>
<td>31.8</td>
<td>10.3</td>
</tr>
<tr>
<td>CMS: Un-Proctored Quiz Group a</td>
<td>39</td>
<td>52.7</td>
<td>76.6</td>
<td>23.9</td>
<td>14.4</td>
</tr>
</tbody>
</table>

a Difference from Section 1 to Section 2 statistically significant (p = 0.01)
b Difference between groups in this column is significant (p = 0.009)

D. Time-on-task

The CMS allowed time-on-task to be tracked using content page hits and practice quiz access. As mentioned, online students were allowed an indefinite number of tries on practice quizzes within each chapter. Figure 2 illustrates the average utilization of practice quizzes per student per chapter by the two CMS groups during the semester. Note references to class quizzes.
Students in the online-proctored-quiz group significantly increased their use of practice quizzes in preparation for quiz 2 (p < 0.001). Students in the online-un-proctored group accessed practice quizzes with the same degree of frequency in preparation for Quiz 2 as they did for Quiz 1 (p = 0.374). In preparation for the first class quiz, there was no statistically significant difference in the average number of practice quizzes taken between the two online groups (proctored and un-proctored). However, in preparation for quiz 2, students in the online-proctored group accessed the practice quizzes almost twice as frequently as students in the online-un-proctored group (p = 0.009).

Students in the online-with-proctored-quizzes group tended toward a larger number of content page “hits” during the semester (mean = 60.34) than those in the un-proctored-quizzes group (mean = 51.46). This difference was not statistically significant (p = 0.182).

Pearson correlations were also run between change score, content hits and practice quiz access. A weak, yet significant correlation (0.257; p = 0.024) was found between improvement in change score and a greater number of “hits” on content pages. A moderate and significant correlation (0.401; p < 0.001) was found between improvement in change score and greater use of practice quizzes. These finding were also checked by dividing the students into two groups based on the mean change score for all CMS users. The results are illustrated in Figure 3. Students in the top half of change score improvement, had significantly more content page “hits” (p = 0.047), and accessed significantly more practice quizzes in preparation for the first class quiz (p = 0.011) and the second class quiz (p = 0.033), as compared to students in the lower half.
V. DISCUSSION

Online delivery of course content was found to be as effective as text-based delivery when paired with proctored testing. In this case effectiveness (learning) was defined by the degree of improvement in post-test score as compared to the pre-test. Student learning trended toward being higher in the online-proctored group as compared to the text-based-proctored group, however, this difference did not reach a level of statistical significance. This finding supports the continued use of online instruction in medical terminology for students in a self-paced, asynchronous format.

Time spent online, as defined by content page “hits,” was weakly correlated with learning, whereas the use of online practice quizzes had a relatively stronger correlation. According to Gilbert’s model, this would be expected, since the practice quizzes more actively engage the students and provide feedback on learning mastery. The practice quizzes were designed to give unlimited access to small, random samplings of the question database for self-testing and match the manner in which competency would be tested during class quizzes. Maintaining a strong relationship between self-testing (practice quizzes) and assessment (exams) appeared to motivate proctored students as evidence by a statistically significant increase in the use of practice quizzes. This supports both the Gilbert and Chickering models that emphasize relevant time-on-task.

Proctored testing proved to be a better facilitator of learning than un-proctored testing, for the online students. Of particular interest was that the online, un-proctored students had the highest average scores of the three groups on their class quizzes, but they also had the lowest performance on learning as defined by pre/post-test change score. Figure 4 illustrates the disparity, for all three groups, between average scores on class quizzes and the pre/post test change score. For both proctored-quiz groups (CMS and book), the scores on the class quizzes correlated significantly (p < 0.05) with pre/post test change scores. This would be expected since the same database of questions was used to prepare the practice quizzes and class quizzes. Although the same random sampling scheme was used to prepare practice and class quizzes for all CMS students, the correlation between scores on the class quizzes and pre/post-test change scores failed to reach a level of significance for the un-proctored group.
To attempt to interpret the difference in performance and behavior, it is necessary to breakdown how synchronous, proctored testing differs from its asynchronous, un-proctored counterpart in this study. These differences center primarily on:

(a) Student supervision: In the proctored group, faculty supervision was present during the examination process.

(b) Structure: In synchronous testing, a “date and time,” combined with a visit to a classroom, provides a degree of structure to the assessment process that is consistent with the experience of most students up to this point in their academic career. This same degree of structure or finality may not exist for a “window” of time and a home or lab computer.

(c) Environment: Even though testing and grading is designed to be very individual (one test, one student) with confidential feedback (grades), it occurs in a communal setting (the classroom). In a sense, synchronous testing may provide a degree of group pressure, motivation or competitiveness that impacts learning.

(d) Intra-test feedback and technical support: The presence of the instructor during the testing process facilitates login, resolution of technical problems and clarification of test questions that may be unclear.

The precise nature of medical terms minimized ambiguity in question construction. This was evident because very few questions arose during the proctored quiz sessions. Most questions that came up were after the quiz was completed and centered on interpretation of the immediate grades and feedback given by the CMS. A few questions regarding grades and CMS feedback were received via email for students in the un-proctored group. There were two reported cases of technical difficulty (“computer lockup”) by students who were taking the quiz un-proctored. In both cases, the quiz was reset with the student
completing it within the specified window. For these reasons, it seems reasonable to rule out the “Intra-
test” feedback effect as meaningful in explaining the difference in performance and behavior between the
two online groups.

Structure and environment effects, as described above, are difficult to analyze in the context of this study
design. However, had students in the un-proctored group been handicapped with respect to these testing
features, poor performance would be expected for both the graded, class quizzes and the pre/post-test
change scores. It should also be noted that a higher average and narrower range for the class quizzes make
these results more difficult to interpret. Greater variability in the class quizzes might have made it easier
to separate and characterize low-change-score students who performed well on in-class quizzes from
those who performed poorly on in-class quizzes.

The level of student supervision may leave some readers to ponder whether academic dishonesty was a
factor in the results seen here. Academic dishonesty is a feature of the traditional classroom setting; and a
number of authors have discussed the question of student supervision in the classroom and how its
absence might affect the online classroom [10–21]. The design employed here is insufficient to answer
this important question. The reader is left to consider these alternatives in light of the results and their
own experience in the classroom.

In conclusion, the online course management system, in combination with the use of proctored quizzes,
was more effective in learning than CMS-based, un-proctored testing. In addition, online course design
should incorporate meaningful time-on-task (e.g. use of self-tests or practice quizzes) to facilitate the
learning process. Finally, these findings support the arguments that instructor supervision is an important
component of online delivery and facilitates its role in learning.

VI. REFERENCES

2. Chickering, A. W., Gamson, Z. F. “Seven Principles for Practice in Undergraduate Education.”
3. Long, D. R. “Time-on-task in Beginning Students of Spanish at the University Level: A Case Study,”
Paper presented at the Annual Meeting of the American Council on the Teaching of Foreign
Languages. ERIC Document Reproduction Service No. ED 239 514. San Francisco, CA, November
1983.
Study of the Feasibility and Utility for National Policy of Instructional ‘Good Practice’ Indicators in
Undergraduate Education.” ERIC Document Reproduction Service No. ED 372 718. Washington,
5. Pintrich, P. R. “A Process-oriented View of Student Motivation and Cognition”. In: Stark, J, Mets,
L. (Eds.), Improving Teaching and Learning Through Research, 65–79. San Francisco, CA: Jossey-
December 2003.

**VII. ABOUT THE AUTHORS**

**Gregory S. Wellman**, Ph.D. is Associate Professor of Pharmacy at Ferris State University. He received his Ph.D. in Pharmaceuticals Administration from The Ohio State University and currently teaches pharmacoconomics, biostatistics and communications in the Doctor of Pharmacy program. Email: wellmang@ferris.edu.

**Henryk Marcinkiewicz** is the Associate Vice President for Academic Affairs at Pennsylvania College of Technology. He was the Founding Director of the award-winning Center for Teaching, Learning, & Faculty Development at Ferris State University in Michigan. He has worked as an educator in a diversity of areas around the world. Marcinkiewicz's doctoral education at Penn State was in instructional systems. His interests include faculty development as well as identifying the conditions under which people adopt innovations, specifically, the use of technology in teaching. Email: hmarcink@pct.edu.