Graduate-Level Science Classes: How Important Are Extra Time, Open Notes and Book When Evaluating Students with Calculation-Oriented Exams?

K.R. Brye[1] and M.C. Savin
Department of Crop, Soil, and Environmental Sciences

E.E. Gbur[2]
Agricultural Statistics Laboratory,

University of Arkansas
Fayetteville, AR 72701

Abstract

Students are known to have individual learning styles and some have preconceived notions, many of which are misconceptions, of the level of difficulty associated with some college-level subjects, particularly in math and the sciences. Some of those preconceived notions may carry through undergraduate education and into graduate school; thus affecting academic performance. Additionally, the added pressure of demonstrating subject mastery under a specific time limit and without means to verify accuracy during testing can produce at least some anxiety in many students. Given that there are typically higher expectations for graduate students, who have experienced three or more years of college and sought more in-depth study, one might expect that fewer barriers to learning new science and demonstrating learned knowledge through written assessments exist with graduate students. The objective of this study was to evaluate how the removal of potential barriers to test performance (i.e., restricted time and few supplemental materials) influenced scores on calculation-oriented exams in a graduate-level Soil Physics class. Results indicated that scores on conceptual-based questions did not significantly improve, but that scores on calculation-based questions, where students were expected to apply basic knowledge to solve problems in a context in which they were somewhat unfamiliar, improved greatly when extra time and the use of class notes and the textbook were allowed for an identical make-up exam. Overall, this study suggests that the combination of extra time and the use of class notes and the textbook on a calculation-based, make-up exam resulted in enhanced student learning by promoting refinement of problem-solving and application skills in the absence of limited-time- and limited-resource-induced stress.

Introduction

Science in general is perceived by many students as difficult subject matter to master (Tobias, 1985). Some of the perceptions associated with learning science may be based on the conveyance during early years of education that some disciplines, like physics, are inherently difficult due to required knowledge of mathematics (Tobias, 1985) or that many facets of learning science require the ability to reason (Piaget, 1964). Similarly, many students lose confidence in their ability to learn science because they choose to learn science by rote memorization (Novak, 1988; Cavallo, et al., 2003), which often leads to misconceptions and misunderstandings (BouJaoude, 1992), rather than obtaining deeper conceptual understanding that allows them to link and apply new ideas with previous knowledge (Zoller, et al., 1997; Cavallo, et al., 2003).

The ability of students to demonstrate learned science knowledge may also be related to the mode of assessment, specifically how examinations are constructed and administered. Logan and Hazel (1999) reported that first-year college students in a physics curriculum performed better on written exams when i) questions were formulated in plain English versus more academic language, ii) questions were formulated without the use of double negatives, iii) diagrams were used in the question or allowed as a response, and iv) abstract questions were used without a context versus questions constructed within a context.

From survey data of Israeli and American college students, mostly in the early years of science-oriented degree programs, Zoller, et al. (1997) concluded that written exams in which time was unlimited and any materials were allowed were preferred over closed-book exams administered in limited time. Some survey respondents indicated that time-unlimited and open-book exams created an atmosphere with

[1]Corresponding author, 115 Plant Sciences Building; Email: kbrye@uark.edu
[2]101 Agricultural Annex

less anxiety and emotional stress so that they were better able to comprehend, think about, and analyze questions, thus leading to an inherent potential for better scores (Zoller, et al., 1997). In contrast, other survey respondents indicated that time-unlimited exams led to a higher probability of second-guessing and that meaningful learning was not actually accomplished because students would search for rather than think about the right answer (Zoller, et al., 1997).

In the context of undergraduate agronomic-science education, Bacon and Beyrouty (1988) indicated that examinations are not necessarily learning experiences because students often fail to attempt to understand why responses are incorrect. Furthermore, Foth (1974) suggested that, if given the opportunity to correct mistakes, learning can be enhanced. Several educators have shown that undergraduates, whether individually or in groups, demonstrated a higher degree of comprehension when provided the opportunity to retake quizzes or exams (Lewis, 1973, 1977; Sorensen, 1977; Longer, et al., 1987; Bacon and Beyrouty, 1988). Clearly, learning and assessment preferences exist at the undergraduate level in science-related subjects, it is less clear to what degree assessment-style preferences or barriers exist, if at all, in science-related subjects at the graduate level where, in theory, lack of academic experience should be less of an issue.

Students in graduate school, regardless of scientific discipline, generally have higher expectations placed upon them by advisors and instructors than were placed on them as undergraduate students. Specifically, advisors and instructors generally expect graduate students to be able to think more critically and independently and be better able to apply their knowledge than undergraduate counterparts because of more experience in the university and academic setting. Though not unreasonable, some advisors and instructors also expect that graduate students be able to assimilate information, exhibit critical and independent thinking, and apply knowledge more rapidly than undergraduates.

Given the choice of graduate students to voluntarily pursue more intensive study within a discipline and previous academic experience as an undergraduate, faculty might expect that fewer barriers to learning new science and demonstrating learned knowledge through written assessments exist with graduate students. However, the stress of test taking does not necessarily diminish with academic experience. In addition, few studies have attempted to differentiate graduate-student performance on conceptual-based compared to calculation-based exam questions in science-oriented subject matter. Therefore, the objective of this study was to evaluate how the combination of extra time and open book/notes influenced scores on calculation-oriented exams in a graduate-level Soil Physics class. We hypothesized that, in general, having extra time and an open book/notes for a calculation-oriented make-up exam would result in greater student learning and appropriate demonstration of learning as a result of higher exam scores. We also hypothesized that if a student answered a question completely correct on the original exam that they would also answer the same question completely correct on the identical make-up exam.

Methods

Background

In Spring 2003, the second of three exams in a graduate-level Soil Physics course taught at the University of Arkansas, covering water movement and solute transport in soil, was administered to eight graduate students, seven of which were Master of Science students and one was a doctoral student. The exam was worth 100 points total, where 25 points were conceptually oriented, short-answer-type questions, and 75 points were calculation-oriented problems. Students were allowed to use one side of a 3-inch x 5-inch note card for formulas only and were given 50 minutes during a regularly scheduled lecture period to complete the exam. Final scores ranged from 39.5 to 84.5 % and averaged 66.5 %. Because of the wide range in scores and several alarmingly low scores, a make-up exam was administered.

Make-up Exam

During the class period following the exam, students were informed of the overall low class performance and that they would be re-taking the same exam. The graded exams were not returned to the students, therefore they were unaware of what responses were correct or incorrect. The make-up exam was exactly the same as the original and was administered 12 days after the original exam. In contrast to the original exam, for the make-up exam, the students were allowed to use their textbook and class notes and were given 2.5 hours during a regularly scheduled laboratory session to complete the exam instead of 50 minutes during their lecture period. Students were informed that they would receive the higher of the point totals per question from either exam so as to not penalize for incorrect answers on the make-up exam that were correct on the original exam.

Despite a rather small sample size upon which to draw statistically significant inference, the circumstances surrounding a study such as this are not ones that any instructor necessarily wants to repeat. However, similar circumstances did arise again the following year in 2004. This time a similar set of data was collected and analyzed regarding the first exam in the same course, which covered basic soil properties and relationships among them.

Repeated Experiment

In 2004, the soil physics class consisted of 10

students, eight M.S. and two Ph.D. students. Similar to that in 2003, the 2004 exam was originally administered during a regularly scheduled 50-minute lecture period and the students were allowed only one side of a 3-inch by 5-inch note card for formulas. However, the percentages of the total possible points for each question type was dominated less by calculation-based questions in 2004 (i.e., 48 % for conceptual and 52 % for calculation-based questions) than in 2003 (i.e., 25 % for conceptual and 75 % for calculation-based questions). In contrast to the format of the make-up exam in 2003, the identical make-up exam in 2004 was administered as a take-home exam, thus students had unlimited time and were able to use any class notes and the text book as needed.

Data Collection and Analysis

The original and make-up exams in 2003 had seven questions, five with multiple parts; for a total of 15 questions that were assigned points separately. For questions with multiple parts graded separately, each part was assumed independent of the other parts within the same question. The original and make-up exams in 2004 also had seven questions, but six of the seven questions had multiple parts for a total of 29 questions that were assigned points separately. The following descriptions of data collected and statistical analyses apply to the original and make-up exams administered in both 2003 and 2004.

Total points earned for each separately graded component of each question were recorded for each student for the original and make-up exams. Individual questions were grouped into two categories of question types: conceptual-based or calculation-based. For purposes of statistical analyses, responses on the original and make-up exam were also grouped into three categories, zero, partial, or full credit, based on the points earned per question relative to the total possible points for that question. Total points earned by individual question and question type (i.e., conceptual versus calculation), and the change in score between the original and make-up exams was then summarized by student and for the whole class. Though the premise of the study conducted and the data collected do not specifically address improved learning, we assumed that a positive change in score was a surrogate result for at least some degree of improved learning.

To ascertain the combined effects of extra time and open notes/book on exam scores and overall class performance, analyses of the frequency distribution of student-question combinations based on earning zero, partial, or full credit on the original and make-up exams for conceptual- and calculation-based questions and for the whole exam were performed (SAS Version 8.1, SAS Institute, Inc., Cary, NC). By collapsing the zero-credit and partial-credit categories into a single partial-credit category, McNemar's test (Fleiss, 1981) was performed by question type and for the whole exam to evaluate whether the probability of obtaining a perfect score or full credit on the original exam was the same as the probability of obtaining a perfect score on the make-up. This procedure allowed us to address the hypothesis that if the student got the question completely right on the original exam that they would also get the question completely right on the make-up exam. In addition, Fisher's exact test (Fleiss, 1981) was performed to determine whether the distribution of students into

Table 1. Class summaries of original and make-up exam scores and the changes in exam score (Make-up - Original) by question type (i.e., conceptual and calculation) and overall for the whole exam from two successive Soil Physics classes taught at the University of Arkansas

	Original Exam				Make-Up Exam				
Year/ Question Type	Low	High	Mean	Ratio of Mean to Maximum Score	Low	High	Mean	Ratio of Mean to Maximum Score	Change in Score
	%								
2003									
Conceptual†	9.0	23.5	20.1	80.4	17.5	25.0	22.3	89.2	2.2
Calculation‡	24.5	62.0	46.4	61.9	47.0	69.5	62.2	82.9	15.8
Whole Exam	39.5	84.5	66.5		67.0	93.5	84.4		17.9
2004									
Conceptual¶	27.5	44.5	38.7	80.6	36.0	48.0	44.4	92.5	5.7
Calculation§	7.5	48.5	33.3	64.0	28.0	52.0	47.4	91.2	14.1
Whole Exam	42.0	85.5	72.0		71.0	100	91.8		17.9

† Maximum score was 25 %.
‡ Maximum score was 75 %.
¶ Maximum score was 48 %.
§ Maximum score was 52 %.

the zero-, partial-, or full-credit categories by question type (conceptual and calculation) and for the whole exam was independent for the original and make-up exams.

Results

Exam Performance

In general, the class performed better on the conceptual- than on the calculation-based questions on the original exam in 2003. However, the conceptual-based questions comprised only 25 % of the total possible points on the exam, while the calculation-based questions comprised the remaining 75 % of the total possible points.

As expected, exam scores by question type and overall for the whole exam increased for the make-up exam in 2003 (Table 1). Scores on the make-up exam increased by 2.2 and 15.8 % on conceptual- and calculation-based questions, respectively, and by 17.9 % overall for the whole exam compared to scores on the original exam. Similar to the original exam, students still tended to performed better on the conceptual- than calculation-based questions despite having more time and their class notes and book as resources to use during the make-up exam.

When examining student performance on individual questions, some interesting results occurred. For 40 of the possible 56 student-conceptual question combinations, scores did not change from the original to the make-up exam in 2003 (Table 2). Of the 40 combinations that did not change, 34 perfect scores were earned on both the original and make-up exams and three had the exact same partial credit earned on both exams (Table 3). However, for three student-conceptual question combinations, two different students for question 2b and one student for question 2d, zero points were earned on the both the original and make-up exams (Tables 2 and 3). Of the conceptual questions on the exam, only one had an overall decrease in mean class score from the original to the make-up exam, but only two of eight students were responsible for the decrease (Table 2).

Of the 64 possible student-calculation question combinations, 25 had scores that did not change from the original to the make-up exam in 2003 (Table 2). Of the 25 calculation-question combinations, 19 perfect scores were earned on both the original and make-up exams (Table 3). In contrast to the conceptual-based questions, no student earned zero points on any calculation-based question on both exams and none of the mean class scores by question decreased from the original to the make-up exam. Interestingly, only one student (Student 8) improved their score from zero credit on one conceptual- and two calculation-based questions on the original exam to full credit or a perfect score on the same questions on the make-up exam (Table 2).

Results of the 2004 make-up exam, with a different set of students and covering different material than that which comprised the 2003 study, were surprisingly similar to the 2003 results. Mean scores on the make-up exam increased on conceptual- and calculation-based questions and overall for the whole exam compared to the original exam with the

Table 2. Summary of 2003 score changes (Make-up - Original) by student for each question type (i.e., conceptual and calculation) and overall for the whole exam

Question Type/Number	Possible Points	% Change in Score Relative to Possible Points by Student								Class Mean
		1	2	3	4	5	6	7	8	
Conceptual										
1a	5	0[†]	0	0	0	60	-20	0	0	5
1b	5	0	0	0	0	80	0	0	0	10
2a	3	0	0	0	0	0	0	0	0	0
2b	3	0	0	0	0	0	50	0	0	6
2c	3	50	0	0	0	50	0	50	50	25
2d	3	33	33	0	0	0	0	0	100[‡]	21
2e	3	33	-33	0	33	0	0	-67	10	-4
Calculation										
3a	5	0	0	0	0	10	10	10	0	4
3b	5	-40	10	20	0	80	40	0	0	14
4a	10	0	0	0	0	30	25	0	0	6.9
4b	10	-35	45	0	0	40	70	60	-20	20
5	10	0	50	50	20	-10	50	0	10	21
6	23	-6	0	81	-15	85	87	0	0	20
7a	5	80	0	10	0	60	60	0	100	39
7b	7	71	71	29	7.1	57	71	0	100	51
Whole Exam	100	5	15	10	0	45	45	6	15	18

[†] A percent change of 0, in most cases, indicates a perfect score on both the original and make-up exam. However, for 3 student-question combinations (2 different students for question 2b and 1 student for question 2d), a 0 % change indicated zero points earned on the both the original and make-up exam.

[‡] A percent change of 100 indicates zero points earned on the original exam and all possible point earned on the make-up exam.

smallest increase occurring with the conceptual-based questions (5.7 %) and the largest increase occurring overall for the whole exam (17.9 %; Table 1). Only one conceptual-based question resulted in an overall decreased mean class score on the make-up exam compared to the original, with only one of 10 students being responsible for the decline (data not shown). Mean class scores on all calculation-based questions increased (data not shown). For the whole exam overall, only six of a possible 290 student-question combinations had zero credit on both the original and make-up exams (Table 4).

Probability of Similar Responses

Based on the frequency distribution, by question type and overall for the whole exam, for student-question combinations earning zero, partial, or full credit (Tables 3 and 4), an evaluation of the probability of similar full-credit responses on both the original and make-up exams was conducted. In 2003, the percentage of full-credit responses to conceptual-based questions was similar for both exams, 66.1 and 78.6 % for the original and make-up exams, respectively. Furthermore, McNemar's test indicated that the probability of earning a perfect score on the original and make-up exams was the same ($p = 0.092$) for conceptual-based questions. However, the percentage of full-credit responses to calculation-based questions was twice as high for the make-up than the original exam. McNemar's test indicated that the probability of earning a perfect score on the original and make-up exams was indeed significantly different ($p < 0.001$) for calculation-based questions. Consequently, the result of significantly different probabilities of earning perfect scores on both exams on calculation-based questions dictated the result of significantly different probabilities ($p < 0.001$) of earning perfect scores on both exams for the whole exam as well. It is also interesting to note that Fisher's exact test indicated that the distribution of student-question combinations into the zero-, partial-, and full-credit categories for the make-up exam was dependent ($p < 0.001$) on the resulting distribution for the original exam for conceptual- and calculation-based questions and overall for the whole exam; thus there appears to be a linkage between performance on the make-up with that on the original exam.

For 2004 exam results, McNemar's test indicated that the probability of obtaining a perfect score was different ($p < 0.001$) for the make-up as compared to the original exam for the conceptual- and calculation-based question sets and overall for the whole exam. This result is in contrast to that from 2003 in which the

Table 3. Frequency distribution of student (n=8)-question (n=15) combinations based on earning zero, partial, or full credit on the original and make-up exams for conceptual-(n =56 observations) and calculation-based (n=64 observations) questions and overall for the whole exam (n=120 total observations) from the 2003 Soil Physics class

Original Exam	Make-Up Exam			Totals
	Zero credit	Partial credit	Full credit	
		Conceptual Questions		
Zero credit	3	2	1	6
Partial credit	0	4	9	13
Full credit	0	3	34	37
Totals	3	9	44	56
		Calculation Questions		
Zero credit	0	3	2	5
Partial credit	0	20	19	39
Full credit	0	1	19	20
Totals	0	24	40	64
		Whole Exam		
Zero credit	3	5	3	11
Partial credit	0	24	28	52
Full credit	0	4	53	57
Totals	3	33	84	120

probability of obtaining a perfect score on both the original and make-up exams for conceptual-based questions was statistically similar. The difference may be a result of more separately graded conceptual-based questions on the 2004 exam (17) than those which were on the 2003 exam (7). In addition, Fisher's exact test indicated that the distribution of student-question combinations into zero-, partial, and full-credit categories from the make-up exam was dependent ($p < 0.001$) on the distribution from the original exam for the concept-based questions and overall for the whole exam, but not for the calculation-based questions ($p = 0.128$). This result is also in contrast to that from 2003, but may be related to overall easier material covered on the 2004 exam compared to that covered on the 2003 exam.

Discussion

Since students were not aware of whether their responses to questions were correct on the original exams, the results of identical make-up exams in a graduate-level, science class indicate that extra time and open notes/book only slightly improved scores on conceptual-based questions. However, significant improvement on calculation-based questions occurred, which resulted in higher overall scores for the whole exam, resulted when extra time and open notes/book were allowed. Regardless of whether the make-up exams were administered with extra time in class or as take-home exams, these results suggest that students were better able to comprehend, analyze, and reason through calculation-based questions that required some level of application to situations they were not exactly familiar with when limited-time- and limited-material-induced stress and pressure were minimized. Consequently, it is likely that, at least with calculation-based questions, learning improved.

Table 4. Frequency distribution of student (n=10)-question (n=29) combinations based on earning zero, partial, or full credit on the original and make-up exams for conceptual-(n =170 observations) and calculation-based (n=120 observations) questions and overall for the whole exam (n=290 total observations) from the 2004 Soil Physics class

Original Exam	Make-Up Exam: Zero credit	Partial credit	Full credit	Totals
		Conceptual Questions		
Zero credit	2	1	10	13
Partial credit	1	13	15	29
Full credit	1	0	127	128
Totals	4	14	152	170
		Calculation Questions		
Zero credit	6	6	25	37
Partial credit	1	12	27	40
Full credit	5	6	32	43
Totals	12	24	84	120
		Whole Exam		
Zero credit	8	7	35	50
Partial credit	2	25	42	69
Full credit	6	6	159	171
Totals	16	38	236	290

Improved scores on make-up exams do not necessarily equate to improved learning. Students may study and prepare differently for the initial and make-up exam. Some may argue that students will study and prepare more for a make-up exam, especially after experiencing difficulty on the initial exam; thus a higher score is more a reflection of being better prepared than how the exam was administered. However, for the purposes of this study, we assumed at the onset that an improved score, whether on a question-by-question basis or overall for the whole exam, represented at least some degree of improved learning.

Although it has been reported that unlimited time may increase the potential for second-guessing (Zoller, et al., 1997), unlimited time and

open book/notes do not necessarily mean that the students did not learn the material. Bacon and Beyrouty (1988) reported that greater than 84 % of Cereal Crop and Introductory Soils undergraduate students surveyed indicated that retaking exams improved their understanding of the material; thus, it can be deduced that learning also was enhanced by the exam-retake procedure.

Also, when open notes/book are allowed, whether for in-class or for take-home exams, that the challenge becomes how well can the students find the right answer rather than how well do they actually know the material themselves (Zoller, et al., 1997). For the calculation-based questions on the two sets of exams presented in this study, finding correct formulas to use from class notes or the textbook was possible and encouraged, but searching for the correct final answer in class notes or in the textbook was not possible since the exams were designed to test the students' ability to apply their knowledge in somewhat unfamiliar contexts. In the case of the subject matter of this study (i.e., soil physics), we feel strongly that the improved scores on both make-up exams reflect the positive effect of additional time for students to synthesize class material to solve problems presented to them in contexts in which they were somewhat unfamiliar. We also feel strongly that the improved scores reflect the positive psychological effect on students of knowing that their performance on the make-up exams would not hinge on how well they could memorize certain class material.

For applied sciences such as soil physics that are heavily calculation- and problem-solving-oriented, the learning challenge is often not with the actual subject matter, but rather the learning challenge arises when students are expected to apply their knowledge to solve problems in a reasonable manner in unfamiliar contexts (Logan and Hazel, 1999). Even graduate students often fail to realize that real-life situations will frequently arise in their future work environment that require application of basic knowledge to problems, questions, or contexts with which they may not be familiar. Therefore, providing the opportunity for students to refine their problem solving and application skills on make-up exams, if the need arises, by allowing them to take more time and use existing resources to address a problem may be a more valuable educational experience than encouraging quick, and often hasty, responses under time-limited examination procedures.

Summary

Two time- and resource-limited exams were re-administered with more time and the allowed use of class notes and the textbook in a graduate-level Soil Physics class. Student performance was recorded by question on the original and make-up exams. Results indicated that scores on conceptual-based questions did not significantly improve, but that scores on calculation-based questions, where students were expected to apply basic knowledge to solve problems in a context in which they were somewhat unfamiliar, improved greatly when extra time and the use of class notes and the textbook were allowed for identical make-up exams. Results also indicated that the probability of obtaining a perfect score on the make-up exams for calculation-based questions was significantly higher than that for the original exams. Overall, this study suggests that the combination of extra time and the use of class notes and the textbook resulted in enhanced student learning from a calculation-based examination in a graduate-level, science course by promoting refinement of problem-solving and application skills.

Literature Cited

Bacon, R.K., & Beyrouty, C.A. 1988. Test retakes by groups of students as a technique to enhance learning. Journal of Agronomic Education, 17(2), 99-101.

BouJaoude, S.B. 1992. The relationship between students' learning, reasoning ability and students' understanding and problem solving of genetics topics. Journal of Research in Science Teaching, 29(7), 687-699.

Cavallo, A., Rozman, M., Blickenstaff, J., & Walker, N. 2003. Learning, reasoning, motivation, and epistemological beliefs. Journal of College Science Teaching, 33(3), 18-23.

Fleiss, J.L. 1981. Statistical methods for rates and proportions. 2nd Ed. John Wiley & Sons, Inc., NY.

Foth, H.D. 1974. Improving learning with mastery learning. National Association of Colleges and Teachers of Agriculture Journal, 18, 3-5.

Lewis, D.T. 1973. A course in soil morphology, classification, and survey-objectives, methods, and student response. Journal of Agronomic Education, 2, 80-84.

Lewis, D.T. 1977. Multiple testing as related to student learning in a course on soil morphology, classification, and survey. Journal of Agronomic Education, 6, 9-13.

Logan, P., & Hazel, E. 1999. Language background and assessment in the physical sciences. Assessment and Evaluation in Higher Education, 24(1), 53-65.

Longer, D.E., Marx, D.B., & Albers, D.W. 1987. The influence of unannounced partial retesting on learning and classroom attendance. Journal of Agronomic Education, 16, 3-5.

Novak, J.D. 1988. Learning science and the science of learning. Studies in Science Education, 15, 77-101.

Piaget, J. 1964. Cognitive development in children: Piaget, development and learning. Journal of Research in Science Teaching, 2(3), 176-180.

Sorensen, R.C. 1977. An analysis of students' grades in a teaching system utilizing repeatable examinations. Journal of Agronomic Education, 6, 51-53.

Tobias, S. 1985. Math anxiety and physics: Some thoughts on learning 'difficult' subjects. Physics Today, June, 61-68.

Zoller, U., Ben-Chaim, D., & Kamm, S.D. 1997. Examination-type preferences of college students and their faculty in Israel and USA: A comparative study. School Science and Mathematics, 97(1), 3-12.