The Impact of Sequential Tests on Student Outcomes: An Exploratory Analysis

Lynn A. Fish

Instructors typically consider material sequencing when developing courses; however, other factors may affect student performance. In this study, student performance when taking quizzes in sequential classes reveals that student performance on quizzes in the second class significantly drops when a quiz is given on the same day in the first class. No effect due the change in quiz sequencing in the first class was observed over the course of the semester. To an instructor, student requests to change the date of a quiz due to another quiz being given on the same date may be a legitimate request, and instructors should carefully consider course requirements sequencing in back-to-back classes.

Keywords: course sequencing; syllabus; testing

INTRODUCTION

Effective planning of instruction is critical to every educational field. Planning incorporates four key aspects: learning theory, critical learning factors in instruction, establishing and sequencing objectives, and selecting appropriate instructional strategies (Stitt-Gohdes, 2002). While the key concept driving instruction is the observable and measurable objectives, they must be ordered in a logical way as appropriate sequencing is critical to effective student learning. In planning, instructors are often faced with decisions regarding course sequencing of curriculum and the required measurement of the objectives through various assignments,
quizzes and examinations. Typically in the college environment, the instructor’s plans are conveyed to the student body through the syllabus. Planning becomes more complicated when addressing scheduling assignments, projects, quizzes and examinations for a common student body that attends classes back-to-back. Student complaints such as “but I have another test in Course X on that date? Professor, can’t you please change the date of ours?” are often heard across campuses. As an instructor should you change the test date? Will it really affect student performance on the test? In this paper, student performance from two classes with a common student body and the sequencing of testing are compared. While a similar study does not exist, as shown in the literature review that follows, one should regard this work as exploratory since other confounding variables, such as the requirements placed on students during the last third of the semester or where the control group students were prior to or following the class of interest, are not tested.

**Literature Review**

Instructor interest and publication in business education has increased in the past decades. A sampling of literature, as shown in Table 1, reveals that instructors have explored various topics within this area; however, a lack of research on business student performance and assignment timing over a semester exists.

Several researchers offer insight into multiple choice sequencing, color of tests to deter cheating, distance learning, exam seat location, exam format, paper-based versus on-line testing, testing frequency, and test time. With respect to multiple choice question sequencing, the recommendations are insightful but confounding. Several researchers cite potentially positive effects on student performance when questions are ordered (reverse or forward) in relation to the course material (Heck & Stout, 1991; Canlar & Jackson, 1991; Carlson & Ostrosky, 1992). However, other researchers found no relationship between various sequencing formats and student performance (Chidomere, 1989; Geiger & Simons, 1994; Peek, 1994; Togo, 2002). To confound the recommendations further, while Heck and
Stout (1991) and Carlson and Ostrosky (1992) found a potential advantage to multiple choice questions ordered chronologically, they conclude that there is no significant effect on students’ mean performance, exam reliability, or item validity. It is also interesting to note that much of the work on multiple-choice sequencing occurred in the last century, with few recent published studies on this topic.

With respect to exam formats, researchers again find differing recommendations. Lawrence and Singhania (2004) found no significant difference in student performance on multiple choice and written exams. However, Holley and Jenkins (1993), and Phillips (1999), note a significant relationship between format and greater cognitive development—that is better students tend to perform better on advanced exam formats, such as short answer and essay. Swartz (2006) explores the impact of self-reported confidence intervals and differing exam formats. His results show that students’ performance with confidence interval inclusion in testing is related to their level of understanding, but these results will require more in-depth analysis by instructors to explore the richness of the information and potential re-writing of exams.

Several studies explore traditional testing environments. Geiger and Simons (1994) found no relationship between student test time, item ordering, and student performance. Recently, Clary, Wandersee and Elias (2007) reported that exam color (as in similar multiple-choice ordered exams with a different paper color) has no significant effect on student performance. They also indicate that exam scores increase as a semester progresses. Kalinowski and Taper (2007) found no effect of seat location on exam performance. Kling, Miller and Reardon (2005) cite increasing testing frequency as positively effecting student performance and higher faculty evaluations.

Distance learning and on-line testing are additional testing areas open for exploration. A sampling of research in this arena includes work by Lawrence and Singhania (2004) and Anakwe (2008). Lawrence and Singhania (2004) found that traditional, in-class students perform better than distance learners on similar exams. While, Anakwe (2008) found no difference in student scores between on-line testing and paper-based testing. In a similar vein, Palocsay and Stevens (2008) explore
the relationship between student performance on web-based homework and tutoring. They found that the results are dependent upon the teacher experience and student academic competence.

Recent studies have emerged to explore optimal sequencing of business curriculum and student course assessment activities. Specifically, Tyler and Tyler (2006) propose recommendations for sequencing activities in ethics based upon the trans-theoretical model of change to assess student readiness. They note that assessment of student activities and assignments is dependent upon the level of learning that the instructor wishes to assess, the students and the educational mission of the institution. Hays, Bousdine-Charmeeva, Goldstein, Hill and Scavarda (2007) explore and propose a framework for grouping and sequencing of topics in operations management. They note that individual student’s learning style such as holistic (or global) or sequential learners should be considered in curriculum sequencing. Recently, Pepper and Pathak (2008) explore the impact of instructor evaluation of student classroom participation in final grade evaluation. They indicate that student perception of fairness is related to explicitness in grading, frequency of feedback, and proactive behavior on the part of the instructor.

While evaluating teaching effectiveness in music instruction, McAllister (2008/2009) notes that in an effective lesson plans instructors should give consideration to the order of performance tasks to increase the likelihood of success. She proposes that instructors should plan for proactive teaching which requires the instructor to purposefully structure effective learning situations in order to elicit correct response, and thereby bringing about increased opportunities for student success. Likewise, instructors, regardless of the field of instruction, should consider the order of performance tasks to increase the likelihood of success. This prompts the question as to whether a student’s claim that dual-activity performance in back-to-back classes may impact upon their performance is legitimate. That is, the hypothesis to be tested is:

H1: The impact of testing in a prior class has no effect upon testing in an immediate following class.
Based upon the sampling cited in Table 1 and discussed here, no business studies exist to test this hypothesis.

**TABLE 1. Research on Business Student Testing Issues.**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Researcher</th>
<th>General Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color of tests to deter cheating</td>
<td>Clary et al., (2007)</td>
<td>Exam color has no significant effect.</td>
</tr>
<tr>
<td>Curriculum sequencing</td>
<td>Tyler and Tyler (2006)</td>
<td>Level of learning that the instructor wishes to assess, the students and the educational mission of the institution</td>
</tr>
<tr>
<td></td>
<td>Hays et al., (2007)</td>
<td>Individual student’s learning style should be considered in curriculum sequencing.</td>
</tr>
<tr>
<td>Distance learning</td>
<td>Lawrence &amp; Singhania (2004)</td>
<td>Traditional students perform better than distance learning.</td>
</tr>
<tr>
<td>Exam seat location</td>
<td>Kalinowski &amp; Taper (2007)</td>
<td>No effect of seat location on exam performance</td>
</tr>
<tr>
<td>Exam format</td>
<td>Holley &amp; Jenkins (1993)</td>
<td>Association between learning style differences and performance with different exam formats (multiple choice, and open-ended).</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Topic</th>
<th>Researcher</th>
<th>General Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students' perceived acceptance on modified exam format, whereby students responded on confidence with particular questions, was equal to with better accuracy than traditional format.</td>
<td>Swartz (2006)</td>
<td>Students’ perceived acceptance on modified exam format, whereby students responded on confidence with particular questions, was equal to with better accuracy than traditional format.</td>
</tr>
<tr>
<td>Homework format</td>
<td>Palocsay &amp; Stevens (2008)</td>
<td>Student performance on web-based homework and tutoring versus traditional dependent upon teacher experience and student academic competence.</td>
</tr>
<tr>
<td>Multiple choice questions sequencing</td>
<td>Chidomere (1989)</td>
<td>Question sequencing had no significant effect on examination performance.</td>
</tr>
<tr>
<td></td>
<td>Geiger &amp; Simons (1994)</td>
<td>Student performance is not significantly impacted by multiple choice tests whereby questions are arranged by inter-topical sequences.</td>
</tr>
<tr>
<td></td>
<td>Peek (1994)</td>
<td>No significant differences between chapter-order sequencing and randomizing.</td>
</tr>
<tr>
<td></td>
<td>Togo (2002)</td>
<td>Students receiving topically-sequence exams scored higher than random version. Also, no significant effects due to advance organizers.</td>
</tr>
<tr>
<td>Paper-based vs. On-line testing</td>
<td>Anakwe (2008)</td>
<td>No difference in student scores between online tests and paper-based tests.</td>
</tr>
<tr>
<td>Testing frequency</td>
<td>Kling, Miller &amp; Reardon (2005)</td>
<td>Increasing the frequency of testing has significant benefits—improved student performance and higher faculty evaluations.</td>
</tr>
<tr>
<td>Test time</td>
<td>Geiger &amp; Simons (1994)</td>
<td>Test time is not affect by item ordering.</td>
</tr>
</tbody>
</table>
COURSES

The courses under review include two required, undergraduate management courses—management science and operations management, at an AACSB-accredited business school in the northeastern United States. Management science topics include an introduction to quantitative modeling, decision theory, forecasting, linear programming, the transportation method, queuing theory and project management. Operations management topics include an introduction to operations management, operations strategy in a global environment, design of goods and services, managing quality and statistical quality control, process analysis, layout strategy, supply chain management, inventory management, aggregate planning, capacity planning, material requirements planning, scheduling and Just-In-Time.

Management students are required to take both the management science and operations management courses. The faculty recommends that both courses be taken simultaneously. However, when a student is behind or chooses not to for a variety of reasons, students must take management science first, and then operations management. Therefore, during the fall semester, student enrollment is mainly college juniors and a sampling of seniors. During the fall semester, junior students typically are enrolled in a human resources management course, statistics II, introduction to finance, management science, and either operations management or a liberal arts course. During the fall semester, seniors who are enrolled in either of these courses are typically enrolled in liberal arts courses. Also, the management science and operations management courses are only offered at one time period. Therefore, the student population’s course schedules are a relatively closed set.

During the semester, the same instructor taught the management science course that met at 8:30–9:45 a.m. on Tuesdays and Thursdays, and the operations management course that followed it from 10:00–11:15 a.m. Student grade evaluation in both courses includes three examinations (25% each) that are non-cumulative and include a homework portion (12%), and quizzes (25%) in which 7 out of 9 quizzes are included
in the student’s final grade. (Note the lowest 2 quiz scores are dropped, and no make-ups are allowed.) Homework assignments consist of a few problems that are graded on a satisfactory (70% correct), unsatisfactory (40–70% correct) and incomplete (0–40%) scale. For the purposes of testing, both courses are divided into thirds, with three homework assignments and three quizzes occurring prior to the examination on the covered material. During the first two-thirds of the semester, material is sequenced such that a homework assignment is due in one class, while the other class had a quiz. During the last third of the semester, homework is due in both classes on the same day, and quizzes are given in both classes on the same day. Did the students perform significantly better, the same or worse in either class when they have another quiz to study for? Examinations are never scheduled on the same date.

**Results**

Nineteen students attended both classes in the fall 2008 semester and formed the back-to-back group in our study (BTB group). Twelve students attended the management science course but did not attend the operations management class that followed and served as the management science control group (MGT control). Nine students attended the operations management class but were not part of the previous management science course and therefore, served as control group for the operations management class (Ops control).

Since it was assumed that each student’s effort was self-guided, the two lowest quizzes in each class were removed from the dataset for both classes. Interestingly, in the management science course, only 1 (3.22%) student had their lowest two scores occur in the last third of the course, 13 (41.94%) students had their lowest two scores occur in the first two-thirds of the course, and the majority of students - 17 (54.84%) had one score occur in the first two-thirds and the other occur in the last third of the course.

In the management science course, the overall course average was 79.22. The nineteen students in the BTB group had an overall course
average of 79.22, while the twelve students’ overall average was 79.02. The difference was insignificant ($p = .472$). As shown in Figure 1, the overall average quiz grade was 8.03 (out of 10; $\sigma^2 = 1.92$) with the average for quizzes 1-6 at 8.02 (\(\sigma^2 = 2.08\)) and an average for quizzes 7–9 at 8.04 (\(\sigma^2 = 1.92\)). The average quiz grade for the BTB group was 8.06 (\(\sigma^2 = 1.85\)) with an average of 8.04 (\(\sigma^2 = 1.93\)) on quizzes 1–6 and an average of 8.09 (\(\sigma^2 = 1.72\)) on quizzes 7–9. For the MGT control group, the average quiz score was 7.99 (\(\sigma^2 = 2.08\)) with the average on quizzes 1–6 at 8.0 (\(\sigma^2 = 2.35\)) and at 7.96 (\(\sigma^2 = 1.59\)) for quizzes 7–9. Analysis of variance revealed that the variance for the BTB and MGT control groups were not significantly different for their overall quiz average ($p = .573$), quizzes 1–6 ($p = .410$), and quizzes 7–9 ($p = .830$).

T-test comparison of the BTB group’s performance for all of the quizzes versus the MGT control group was insignificant ($p = .714$). Similarly, t-test comparison of the BTB and MGT control groups for the first six quizzes ($p = .875$) and the last three quizzes ($p = .669$) were not significant. Thus, performance on quizzes between the BTB group and the MGT control group was not significantly different.

**FIGURE 1.** Comparison of First Class Averages Q 1–6 and Q 7–9.
control group was not significantly different for the first class taught each day.

For the BTB group, analysis of variance for the management science course revealed that their score on the first six quizzes versus the last three quizzes was insignificant ($p = .682$). Similarly, for the MGT control group, their performance on the first six quizzes was not significantly different than the last three quizzes ($p = .258$). Paired t-test comparison for the BTB group revealed an insignificant difference ($p = .827$) between quizzes 1–6 and quizzes 7–9. For the MGT control group, paired t-test comparison of the first six quizzes versus the last three quizzes was also insignificant ($p = .927$). Thus, for the BTB and MGT control groups, performance on quizzes did not significantly change over the course of the semester.

As previously discussed, the operations management class was divided into the back-to-back (BTB) group and the control group that did not attend the first class (OPS Control group). Interestingly, only 7 (25%) students had their two low scores occur in the last third, 4 (14.3%) students had their low scores occur in the first two-thirds, and the majority (again!) - 17 (60.7%) had one poor grade in the first two-thirds of the course and one in the last third.

For the operations management class, the overall course average was 79.93. The nineteen students in the BTB group had an overall course average of 78.93, and the nine Ops control students’ overall average was 82.05. Again, the overall average is insignificant between the two groups ($p = .225$). As shown in Figure 2, the student’s average quiz grade was 7.81 ($\sigma^2 = 2.46$), with an average of 7.80 ($\sigma^2 = 2.60$) on the first six quizzes and an average of 7.84 ($\sigma^2 = 2.11$) on the last three quizzes. The quiz average for the BTB group was 7.72 ($\sigma^2 = 2.45$) with an average of 7.78 ($\sigma^2 = 2.55$) on quizzes 1–6 and 7.55 ($\sigma^2 = 2.22$) on quizzes 7–9. As for the Ops control group, their overall average was 8.01 ($\sigma^2 = 2.45$) with an average of 7.84 ($\sigma^2 = 2.77$) on quizzes 1–6 and 8.47 ($\sigma^2 = 1.38$) on quizzes 7–9. Analysis of variance revealed that the BTB and Ops control group were not significantly different with respect to overall quiz average ($p = .979$), quizzes 1–6 ($p = .713$), and quizzes 7–9 ($p = .318$).
Overall student performance between the BTB and the OPS control groups revealed no significant difference in overall quiz performance between the two groups (p = .224). Students’ performance on the first six quizzes between the BTB group and the OPS control group was insignificant (p = .839). However, the difference in quiz performance on the last three quizzes—quizzes 7, 8 and 9, were significantly different between the groups (p = .03). Thus, for the second class offered in the back-to-back sequence, changing the quiz to the same day had an effect! While some readers may be concerned over the small sample size and its associated bias, the purpose of statistical power calculations is to get a large enough sample without getting a sample that is needlessly large. Cohen (1977) indicates that a small sample is quite adequate where risk of Type I Error is set at p < .05 and the expected degree of relationship is large, as in our case.

For the operations management class, paired t-test comparison revealed an insignificant difference in performance between the first six quizzes and the last three quizzes for the BTB group (p = .477) and the Ops control
group \((p = .361)\). Similarly, analysis of variance revealed an insignificant difference in variation for each set of quizzes for the BTB group \((p = .659)\) and the Ops control group \((p = .133)\). Thus, for each student, their performance in the operations management class did not significantly change across the semester.

At the end of the semester in the second class, the nineteen students in both classes were surveyed regarding their preference regarding quizzes and homework assignments sequencing in the two classes. Twelve of the nineteen students \((63.2\%)\) preferred to have homework due in one class with a quiz in the other, that is, to have quizzes on opposite days. Seven students \((36.8\%)\) preferred to have homework due on the same day with quizzes on the same day. One student did not have a preference.

**Discussion**

The analysis reveals that for the first class, there was no significant difference in performance over the semester; however, the performance in the second class significantly changed. For the first class of the day, student performance was not impacted by whether they had a quiz or project or test in the class immediately following or not. However, this was not the case for the second class. The BTB quiz average dropped, while the Ops control group average quiz performance improved on the last three quizzes. These results imply that student performance is significantly impacted upon by the *immediately prior* class testing. To an instructor with control over testing, the results indicate that the student complaint “But I have another quiz on that day? Can’t you please change the date?” may actually be a legitimate request—particularly for later classes. The fact that the test results were significant with a small sample size is actually more indicative of the power of these results (Cohen, 1977). Students who had two quizzes back-to-back performed worse on the second quiz in the day. However, it should be noted that where the other control students originated from, another course or even home, is unknown. The control group in the operations management class was under the same environment throughout the semester, and in fact, their
results improved over the last third—particularly in comparison to the back-to-back group.

It is interesting to note that student feedback mirrored the quantitative results, that is, the preference for opposite testing dates was generally preferred from a performance and a qualitative viewpoint.

Therefore, in creating the course schedule, when an instructor teaches back-to-back classes with a common student body, the instructor should consider quiz timing over the course of a semester as it can have a detrimental effect upon student performance. The implied recommendations from this study are that instructors should not have testing on the same day as student performance in the second class will suffer.

This study has a number of failings. First, the number of students in the dataset is low. This fact alone will raise researcher’s concerns. However, since significance was noted with such a small dataset, this in fact, strengthens the argument that there is a difference between the groups. Second, the study assumes the control groups did not come from or go to another class with testing in the immediately preceding or prior class. Third, two-thirds of the quizzes occurred on different days while only one-third of the quizzes occurred on the same day. Thus the dataset was unequal and in the future the study should consider a dataset with equal treatments.

REFERENCES


