

Exploring Variation in Student Academic Performance: Can Achievement in an Immersive Case Study Project Predict Exam Score in an Introductory Accounting Course?

Penelope M. Lyman

University of North Georgia
penelope.lyman@ung.edu

Alexander E. Olvido

University of North Georgia
alex.olvido@ung.edu

Abstract: In modern university education, quantitative analytical skills seem best acquired through deep learning of complex, multi-faceted problems. Our quasi-experimental design tested whether student achievement in an immersive classroom case study might affect subsequent academic performance, presumably reflecting deeper learning of fundamental principles in an accounting course. We analyzed exam scores of three behavior-based student groups: (a) “OOP,” who Opted Out of the immersive case study Project, (b) “BMP,” who earned Below Median marks on the Project, and (c) “AMP,” who scored At least the Median on the Project. Results indicate that student academic performance declined at effectively equal rates among the three student groups in any given semester. Surprisingly, students’ self-reported deep strategy at the start of the school term more strongly predicted their academic performance, accounting for more than 30% of exam score variation; group membership explained only 1.93% of exam score variation. These results underscore the need to document student learning approaches explicitly in order to complement observations of student classroom behaviors and academic performance.

Keywords: accounting cycle, student approach to learning (SAL), repeated-measures ANOVA

Introduction

Case studies are common ways to engage students in complex learning. A case study can challenge students to reflect expansively and profoundly on a particular subject (Healy & McCutcheon, 2010). However, a case study might overwhelm students conceptually and/or procedurally, especially when illustrating a fundamental organizing principle (e.g., accounting cycle) all at once. For example, Phillips and Heiser (2011) found no difference in exam performance between students who were presented the accounting cycle piecemeal, focusing on “transactions that affect only balance sheet accounts” (p. 683), and students who were presented a slightly more complete accounting cycle early in the course, focusing on “transactions that affect both balance sheet and income statement accounts” (p. 683). Given that scholars have not addressed student learning of the complete accounting cycle (i.e., from analyzing transactions through production of financial statements), we decided to investigate how current pedagogical practices in an introductory accounting course might benefit university students. One such student-centered practice is instructor-guided immersion using two consecutive accounting cycles of a hypothetical consulting company.

Evolution of Student-Centered Classrooms

Higher education in the 20th and 21st centuries has undergone extensive change. Kolb and Kolb (2005) traced the origins of experiential learning theory to John Dewey's pre-WWI advocacy of learning through personal and transformative experiences followed by self-reflection, an idea that opposes the traditional lecturer-centered approach of transmitting "preexisting fixed ideas . . . to the learner" (p. 194). In the United States, the proliferation of similar, self-reflective learning theories and practices was facilitated by urbanization in the first half of the twentieth century (Greenfield, 2013). Post-conflict social benefits (e.g., "U.S. Servicemen's Readjustment Act of 1944," or "the G.I. Bill"; Greenberg, 2011) expanded American higher education programs to accommodate an unprecedented influx of students, fueling experimental research on student learning (McKeachie, 1990). From 1950 to 1980, educational researchers accumulated data to support the idea that across various fields of study, (a) small class sizes and (b) discussion-driven (vs. lecture-dominated) classrooms resulted in greater long-term knowledge retention (McKeachie, 1990). Pelaez et al. (2015) described the specific classroom methods of Samuel Postlethwait, a pioneering researcher and practitioner of student-centered pedagogy in the 1970s.

The growing masses of university students in the decades following WWI sparked concerns about educational quality in a post-industrial world. Two higher education reports in the 1980s, *A Nation at Risk* and *Involvement in Learning*, motivated U.S. college and university leaders to consider student-centered measures of institutional success (e.g., student engagement and learning gains) to complement their traditional focus on reputation-building and resource acquisition (Koljatic & Kuh, 2001). Empirical support for student-centered learning has since continued. Jones and Fields (2001) found that, regardless of academic background, students in an introductory accounting course who had voluntarily participated in "Supplemental Instruction" (SI) improved their test performance. Oddly, Jones and Fields also suggested mandating SI participation by students rather than incorporating the active learning pedagogy of SI tutors into their own lecture-driven courses. In researching active vs. passive teaching styles, Michel et al. (2009) found that active teaching improved specific (i.e., declarative and factual) knowledge but not broad (i.e., synthetic and application-focused) knowledge. Slavich and Zimbardo (2012) lamented the ongoing dominance of traditional lecturing when the reviewed evidence suggested its relative ineffectiveness in student learning.

Beyond U.S. borders, education reformers witnessed a delayed shift towards student-centered pedagogies. For example, until the second Hadow report of 1931, English educational leaders expressed skepticism over Dewey's view of education as a means for advancing society: "In the eyes of many of its citizens, . . . Britain did not require social progress, for it was already a fully developed civilization" (Darling & Nisbet, 2000, p. 45). In contrast, Canadian ministers of education were open to Dewey for practical reasons (Christou, 2013a, 2013b). Responding to the swell of students during economic depression, Ontario's education minister advocated, in his 1932 annual report, for flexible school programs and class schedules that "provide more 'choice' of subjects and 'freedom' for the learner" (Christou, 2013b, p. 572). In Hamburg, a few German schools experimented with progressive curricula immediately after WWI, only to return to more authoritarian teaching with the rise of Nazism (Roith, 2014). In contrast to the militaristic patriotism in neighboring Germany at the time, Dutch educators agreed with Dewey-esque self-actualization in order to prevent children in the Netherlands from adopting blind obedience to authority; however, Dutch education policy never fully embraced Dewey's principles beyond optional "handicraft" classes due to various social and political factors (Stolk et al., 2014, p. 704). Post-WWII rebuilding efforts across Europe eventually led to school reforms that encouraged student-centered teaching and learning (Darling & Nisbet, 2000; Allemann-Ghionda, 2000; Schneider, 2000).

By the 1970s, in order to increase efficiency of learning, educational researchers shifted attention from teacher improvement to understanding the cognitive processes that students undergo. McKeachie (1990), Duff and McKinstry (2007), and Entwistle (2015) reviewed these developments in fascinating detail. Briefly, Marton and Säljö (1976a, 1976b) noted a sharp contrast in how students processed a reading assignment. Some made little effort to find meaning beyond the printed words, while others sought conceptual connection with prior knowledge—the former were identified as using a “surface approach,” the latter as using a “deep approach” (Marton, 1976; Svensson, 1976; Richardson, 2015). British researchers were subsequently able to generalize this reproducible learning dichotomy across various experimental tasks and developed a number of survey instruments to measure these approaches. In addition to standardizing terminology, these survey instruments clarified certain student behaviors associated with surface vs. deep learning approaches.

Hence, the consensus view of modern university education seems to downplay rote memorization of lectured facts and procedures in favor of more interactive, synthesis-driven, and integrative approaches to learning—in short, to foster cognitively deep learning (Hall et al., 2004; Nelson Laird et al., 2014). With the goal of systematic understanding of inherent processes, deep learners characteristically self-reflect often and attempt to make conceptual connections to solve problems. In the classroom, deep learning might take place through interactive discussion of ideas that challenge longstanding beliefs (i.e., “cognitive conflict”; Sargent & Borthick, 2013). Similarly, at program and institutional levels, some “high impact” educational practices that promote student self-reflection include a combination of first-year seminars, cross-disciplinary learning communities, study-abroad opportunities, and writing-intensive courses (Wawrzynski & Baldwin, 2014, p. 56).

Arguably, the “Presage, Process, and Product” (Dunkin & Biddle, 1974; reviewed in Groccia, 2012), or “3-P,” model (Biggs, 1979) provides the clearest framework for operationalizing deep learning. As depicted in Duff and McKinstry (2007), Nelson-Laird et al. (2014), and Duff and Mladenovic (2015), the 3-P model holds that in the “Presage” stage, a student’s personal attributes (e.g., earlier experiences that shaped learning interests) and current learning environment (e.g., physical, curricular, institutional, and social factors) interact to shape student perceptions of specific learning tasks, perceptions that might not always align with those of the instructor. In the “Process” stage, depending on task perception (e.g., easy, challenging, or impossibly difficult to relate to), the student can choose to adopt surface or deep approaches, or some combination of both. Presage factors that cumulatively fail to align with process variables result in low-quality learning outcomes in the “Products” stage. Therefore, an educator can best encourage deep learning in students by aligning presage factors that promote deep learning (e.g., relating course material to students’ personal lives, providing prompt feedback to students, etc.) with deep-learning process variables (e.g., immersive classroom demonstrations) to yield the desired product (e.g., commendable exam scores).

Focus on Accounting Education

Students majoring in business comprise a significant portion of the American population attending university (see Table 318.20, National Center for Education Statistics, U.S. Department of Education, <https://nces.ed.gov>). From 1980 to 2012, the total number of degrees granted by U.S. baccalaureate institutions nearly doubled, growing from 935,140 to more than 1.79 million. In that same time period, the percentage of students graduating with a degree in business (separate from public administration and from legal studies) ranged from 19% to 24%. At the same time, baccalaureate degrees in the social and behavioral sciences, including anecdotally popular fields of psychology, social science, and history, actually declined from 23% to 16%. In other words, approximately one out of every five university students in the United States today has, is, or will be enrolled in an introductory accounting course.

A fundamental framework for learning accounting principles is the accounting cycle. The accounting cycle is a series of steps grounded in accounting “best practices” for organizing financial information into a standardized format. More specifically, the accounting cycle includes (a) analyzing and recording financial transactions in journal entries, (b) posting those journal entries to a ledger that allows preparation of an unadjusted trial balance, (c) journalizing and posting adjusting entries to prepare an adjusted trial balance, (d) preparing a discretionary end-of-period worksheet to analyze and summarize the data, (e) preparing the financial statements, (f) journalizing and posting closing entries, and (g) preparing a post-closing trial balance. Warren et al. (2014) provided a detailed description of the accounting cycle, along with many helpful examples.

The conceptual breadth and depth of the accounting cycle mandates thoughtful analysis during instruction. The accounting cycle requires students to master and synthesize many interrelated concepts, each usually overviewed separately using various illustrative examples. Unfortunately, the seemingly fragmented presentation of the accounting cycle in traditional accounting instruction (e.g., conceptual “scaffolding”; Phillips & Heiser, 2011) might also reinforce a fragmentary understanding of this inherently recursive process. Engendering a deeper understanding of the accounting cycle might, instead, require greater emphasis on system-wide reflection, one that immerses students in balancing a complete set of business transactions for a single-company case study.

Research Objectives and Hypothesis

Here, we documented the effect of voluntary participation by students in an immersive classroom case study. Our findings are important for three reasons: (a) they extend discussion of how best to present the accounting cycle, a fundamental conceptual framework for understanding accounting “best practices”; (b) they invite accounting educators to confront aspects of their teaching that might discourage naïve students from pursuing accounting careers after exposure to realistic scenarios of accounting work; and (c) they suggest how educators within as well as outside the accounting discipline might apply powerful but rarely used statistical tools to measure the effectiveness of case studies in student learning.

By definition, students who adopt deep approaches to learning the accounting cycle will eventually possess superior understanding of its many interrelated facets. Deep learners presumably learn at different rates; as a result, they are likely to demonstrate content mastery at different times during a school term. By taking careful account of as many quantifiable contributing factors as possible—including student learning approach—we can isolate the effect of voluntary participation in an immersive case study about the accounting cycle on student exam performance. More specifically, assuming classroom behaviors indicate the learning strategy that students adopt (i.e., surface approach *versus* deep approach), we hypothesized that level of voluntary participation in an immersive case study about the accounting cycle would predict student exam scores.

Methods

Ethics Statement

The current project was reviewed by the university’s Institutional Review Board (IRB), which exempted the project from full review. We conducted the study in an acceptable educational setting and involved normal educational practices in compliance with subsection 45 CFR 46.101(b)(1) of the U.S. Federal Policy for the Protection of Human Subjects. Student records that we accessed, while confidential, were not highly classified by school administrators and, thus, were appropriate for reporting in the aggregate to guarantee student anonymity. Statistical analyses were performed only

after the end of the school term in which data had been collected, precluding any influence of the current research findings on student course performance. In addition, both authors completed online CITI¹ Training Program on Information Privacy Security and Conflict of Interest.

Course Description

We collected student data from class rosters of Accounting Principles I courses from Summer 2017 through Spring 2019 at a southeastern U.S. university. Successful completion of this course required students (a) to demonstrate an understanding of how accounting generally “works” in any business operation, (b) to analyze, record, and summarize transactions related to financing and operating both service and merchandising businesses, and (c) to prepare and interpret basic financial statements. Readers interested in a lucid description of topical coverage in foundational accounting courses should consult Saudagaran (1996) and Warren and Young (2012). Both Accounting Principles I and its immediate follow-up, Accounting Principles II, ran eight weeks as back-to-back courses in the regular academic calendar (i.e., Fall and Spring semesters), and only five weeks back-to-back during Summer semesters. Typically, students complete Accounting Principles I and II by the end of their second year of university study.

Classroom Setting

A common mode of university classroom instruction emphasizes student collaboration after brief topical lectures. For example, the course instructor might introduce a sample problem with background information, demonstrate a solution to that problem, and then invite students to deliberate among themselves on possible solutions to a different, follow-up problem before the instructor reveals a solution. Typically, students direct questions at the course instructor at any time during the class period, and the instructor might respond forthrightly or with leading questions, invoking the Socratic method (e.g., preparing a case study for classroom presentation; Guess, 2014). In Accounting Principles I, student questions are predominantly procedural (e.g., how to calculate the maturity value of a note receivable) and sometimes conceptual (e.g., how omitting an adjusting entry might impact a financial statement).

Study Population and Focal Assignment

We classified student subjects according to their behavioral response to a voluntary assignment. To facilitate the exploratory nature of the current research, we limited the study population to students enrolled in a single instructor’s course sections. The assignment was a comprehensive case-study problem involving a single hypothetical company. All financial transactions and procedural instructions were given in the course textbook (Warren et al., 2014). Though the project was materially based on widely available textbook content, the optional nature and pedagogical presentation of the project—hereafter, the Optional Cumulative Project (OCP)—afforded each student equal opportunity to learn fundamental accounting practices in a relatively low-stress environment. Students were also allowed to opt out of OCP instruction (i.e., leave class early or not attend at all during OCP instruction).

OCP instruction began during the class session immediately preceding the session when students completed the first exam. The first task of OCP was to enter journal entries on accounting paper by hand in the second month of the new (hypothetical) company’s business transactions, a task

¹ Collaborative IRB Training Initiative

that required them to transfer the post-closing balance from the first month's accounting cycle to the second. Students attempted the bulk of subsequent OCP tasks alongside the instructor in class and were required to complete unfinished tasks as homework (e.g., finish posting to the ledger and prepare the unadjusted, adjusted, and post-closing trial balances). Earning full credit required students to submit all required OCP tasks and correspondingly accurate documentation before the start of the second exam (i.e., one week later in most cases). Students who chose not to participate (i.e., failed to submit any OCP work when due or submitted only a partially completed OCP assignment) were identified as having *Opted Out of the Project* (i.e., "OOP" students). Students who voluntarily completed the OCP but earned *Below Median* marks on the *Project* were identified as "BMP" students, while those who achieved *At least the Median* score on the *Project* comprised the "AMP" student group.

In terms of assignment weight, OCP was virtually inconsequential to final grades. The maximum number of points from required course work during any school term was 569 points such that completion of OCP over three consecutive class periods (out of 21 instructional days in each academic term) determined less than 2% (i.e., 10/569) of the final course grade.

Our definition of student treatment groups based solely on observable behavioral response to a completely voluntary classroom exercise might provide insight into student learning motivation. Choy et al. (2012) argued that students should manifest their learning approaches as observable classroom behaviors before such learning approaches can reliably predict academic outcomes. Because OOP students failed to submit a completed OCP prior to Exam 2, we were not able to determine the primary motivation or reasoning for their non-participation. BMP students completed a low-quality OCP prior to Exam 2, a behavior that, nevertheless, suggests higher learning motivation than OOP students. AMP students completed a high-quality OCP prior to Exam 2, suggesting higher learning motivation than either BMP or OOP students.

Survey on Student Approaches to Learning

Within the first week of each academic term, we surveyed the general learning approaches of enrolled students. The survey instrument—Biggs' Revised Two-Factor Study Process Questionnaire, or RSPQ2F (Biggs et al., 2001)—was a brief series of statements to which participants rated their agreement. Students submitted both a formal consent form and the completed RSPQ2F with only minimal incentive (i.e., extra-credit points amounting to less than 1% of all points from scheduled course activities [= 5/569]).

Earlier studies suggest that RSPQ2F adequately measures the task perception of students who adopt surface *versus* deep(er) approaches to learning in general (Justicia et al., 2008; Lake et al., 2017). While scholars have translated it for non-native English students (e.g., Bati et al., 2010; Stes et al., 2013; Mirghani et al., 2014; Zakariya et al., 2020), the instrument's applicability to students beyond the Hong Kong study population highlighted in Biggs et al. (2001) is questionable (Immekus & Imbrie, 2010). Moreover, Vaughan (2018) argued that Biggs' second-order factors distinguishing learning motivation from learning strategy were context-dependent.

Despite these psychometric disagreements, the conciseness and lexical clarity of the survey items on RSPQ2F are strong. Students and teachers understand that strong agreement with statements such as "I only study seriously what's given out in class or in the course outlines," "I learn some things by rote, going over and over them until I know them by heart even if I do not understand them," or "I believe that lecturers shouldn't expect students to spend significant amounts of time studying material everyone knows won't be examined" corresponds to a learning strategy that promotes superficial understanding of the course material. Likewise, strong agreement with statements such as "I find that I have to do enough work on a topic so that I can form my own conclusions before I am

satisfied,” “I find most new topics interesting and often spend extra time trying to obtain more information about them,” or “I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes” reflects a learning strategy for knowledge retention and deep understanding (e.g., accounting principles and practices).

Student Academic Performance and Background Data

We initially considered analyzing two measures of student academic performance—final course grade and individual exam scores. Constituting multiple observations of the same student across time (i.e., time series), individual exam scores provide statistical and logistical advantages over a single final grade (Neter et al., 1990). Likewise, final course grades mask the effects of putative causal factors (e.g., individual student nervousness or misunderstanding of idiomatic written English) that seem better reflected in individual exam scores. We eventually chose exam scores as the sole measure of student academic performance because they essentially determine final course grade and, unlike final grades, provide repeated observations that allow students to act as their own control group. The latter reason is particularly important in quasi-experimental designs for detecting nuanced performance differences between pre-treatment and post-treatment observations (Thyer, 2012).

We included in our formal, statistical analysis two key items from archived student records: gender and academic load per semester. We deduced gender from the ‘Mr.’ or ‘Ms.’ designation in student records, while academic load was simply the total number of enrolled credit hours at the start of each semester. Notably, summer-enrolled students tend to carry lower academic loads but are challenged with an accelerated instructional and assessment schedule.

Quantitative Analysis of Exam Scores

We borrowed statistical analytical procedures akin to those successfully implemented in animal behavior research (Olvido & Wagner, 2004; Olvido et al., 2010). First, to minimize confounding effects of heterogeneous variance (Sokal & Rohlf, 1981), we transformed all students’ exam scores (originally expressed as percentages) to normalized ranks, thus allowing subsequent statistical tests to reach asymptotic relative efficiencies (~maximum statistical power) equal to or greater than those of any parametric or non-parametric test (Conover, 1999). Also, unlike simple ranks, normalized ranks facilitate unbiased tests of interactions (Conover, 1999).

Then, we applied the following repeated-measures analysis of variance (ANOVA) model:

$$Y_{ijklmn} = \mu_{\dots\dots} + EXAM_i + TERM_j + GEND_{k(j)} + PROJ_{l(jk)} + LOAD_{m(jkl)} + DEEP_{n(jklm)} \\ + (EXAM \times TERM)_{ij} + (EXAM \times GEND)_{ik(j)} + (EXAM \times PROJ)_{il(jk)} \\ + (EXAM \times LOAD)_{im(jkl)} + (EXAM \times DEEP)_{in(jklm)} + \varepsilon_{(ijklmn)},$$

where Y_{ijklmn} is normalized-rank score on the i th exam earned by a student in the j th academic term, of the k th gender and l th OCP treatment group, with m th academic load and n th deep strategy score; $\mu_{\dots\dots}$ and $\varepsilon_{(ijklmn)}$ are model terms for, respectively, the common mean and unexplained variance (= error); $EXAM_i$ denotes the repeated-factor and fixed-effect model term for within-student variation among exam scores ($i = 1, \dots, 7$)—we excluded students who did not participate in all seven exams; $TERM_j$ is the fixed-effect model term denoting variation in exam scores among Summer, Fall, and Spring academic terms ($j = 1, \dots, 6$); $GEND_{k(j)}$ is the random-effect model term denoting variation in exam scores due to student gender ($k = 1, 2$) and nested within academic term; $PROJ_{l(jk)}$ is the random-effect model term denoting variation in exam scores among OCP student groups (i.e., AMP,

BMP, and OOP; $l = 1, 2, 3$) and nested within academic term and student gender; $LOAD_{m(jkl)}$ is the random-effect model term denoting variation in exam scores due to per-semester academic load ($m = 1, \dots, 13$), which varied from 3 to 20 credit-hours, and nested within academic term, student gender, and OCP student group; and $DEEP_{n(jklm)}$ is the random-effect model term denoting variation in exam score due to student self-reported deep strategy score ($n = 1, \dots, 14$), which varied from 10 to 23 in our student sample, and nested within academic term, student gender, OCP student group, and student academic load. Because we measured individual student exam performance only once in each examination period, the ANOVA model's error term not only remained unrepeated but also conflated with variation due to the two-way interaction between exam number and deep-strategy score group (i.e., an error-inflated $(EXAM \times DEEP)_{in(jklm)}$ model term). The lack of within-subject, within-treatment replication only renders the corresponding F tests conservative and, thus, more challenged with detecting statistically significant variation (Neter et al., 1990).

Third, to describe the relative contribution of the ANOVA model terms to total variation in exam scores, we used restricted-maximum likelihood (ReML) estimation procedures to quantify variance components. In requiring independence of exam scores (i.e., uncorrelated observations), ReML estimation of variance components cannot appropriately evaluate statistical significance (unlike repeated-measures ANOVA) and only quantitatively describes the importance of each ANOVA model term to student exam performance.

Finally, we performed all statistical computations and summaries using Windows-based computer applications. For most descriptive statistics and calculation of all F-test mean-squares and ReML-based variances, we used the MEANS, GLM, and VARCOMP procedures, respectively, in SAS 9.3 (SAS Institute, 1999), with final P values obtained online (Soper, 2018). Organizing the various results into more reader-friendly format, we used Apache OpenOffice programs (Version 4.1.2; The Apache Software Foundation, 2014) to generate all figures and summary tables.

Results

Sources of Variation in Student Exam Scores

For any given student, exam scores significantly varied across the course (see Line 1, Table 1). Mean exam scores appeared to decline across the seven exams (see Figure 1). The interaction of exam number with other variance factors (see Lines 7–10, Table 1), including OCP student group (see Line 9, Table 1) was not statistically significant, indicating that exam scores generally declined at equal rates over time across the three OCP student groups (see Figure 1).

Table 1. Repeated-measures analysis of variance on normalized-rank student exam scores from an introductory accounting course (Summer 2017 through Spring 2019)

Line	Variance Model Term	df [§]	Type III	
			Mean Squares	Observed F
1	Exam number, EXAM (repeated factor)	6	8.375	17.91 ***
2	Academic term, TERM	5	6.627	2.30
3	Student gender, GEND, nested within TERM	6	2.881	0.64
4	Optional cumulative project student group, PROJ, nested within TERM and GEND	18	4.522	1.34

5	Academic load, LOAD, nested within TERM, GEND, and PROJ	51	3.371	1.21
6	Deep strategy score, DEEP, nested within TERM, GEND, PROJ, and LOAD	41	2.795	6.03 ***
7	EXAM × TERM interaction	30	0.618	1.32
8	EXAM × GEND interaction, nested within TERM	36	0.468	0.90
9	EXAM × PROJ interaction, nested within TERM and GEND	108	0.517	0.97
10	EXAM × LOAD interaction, nested within TERM, GEND, and PROJ	306	0.534	1.15
11	EXAM × DEEP interaction, nested within TERM, GEND, PROJ, and LOAD	323	0.463	---

Note. $N = 133$ students \times 7 exams = 931 observations. Due to the nature of a repeated-measures design, the last interaction term (Line 11) cannot be evaluated separately from unexplained (= error) variance. §degrees of freedom. *** $p < 0.001$.

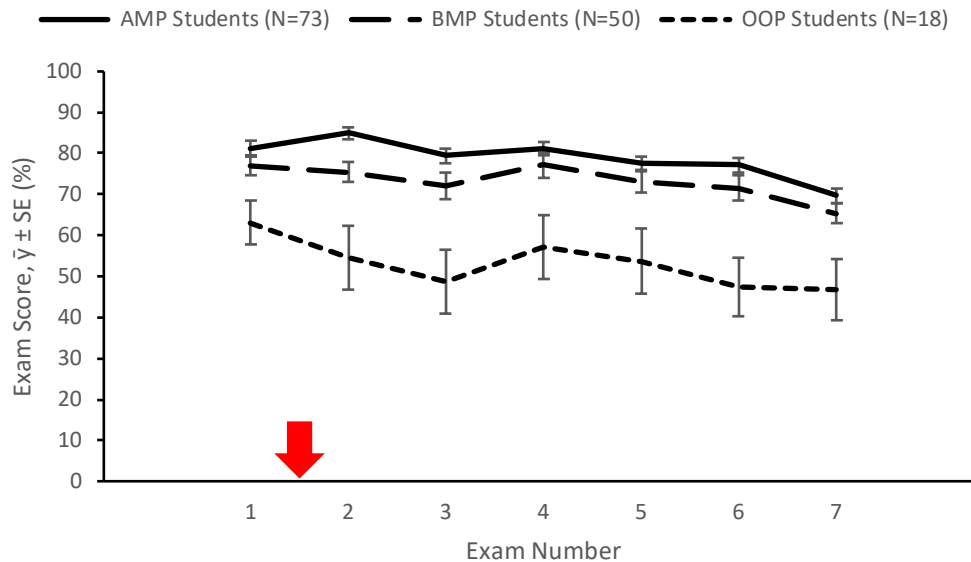


Figure 1. Variation in student exam performance of three student groups in an introductory accounting course (Summer 2017 through Spring 2019). Red arrow demarcates end of immersive OCP instruction featuring an overview of the accounting cycle.

Exam scores significantly varied across deep strategy scores (see Line 6, Table 1). Students who self-reported lower and higher deep strategies tended to achieve the highest exam scores (see Figure 2). The lowest exam scores (i.e., cumulative final exam) appeared to correspond to those students who had self-reported a deep strategy score of 18 out of a maximum 25 on the Biggs RSPQ2F survey instrument.

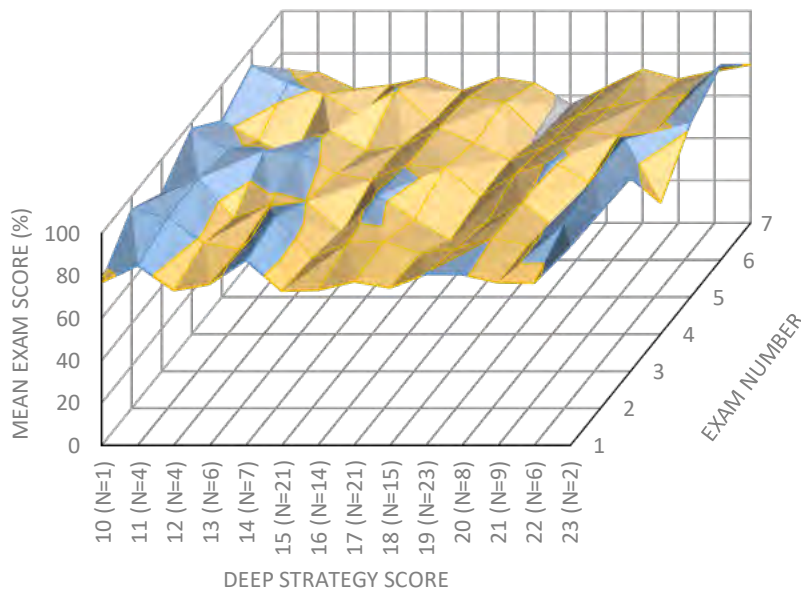


Figure 2. Variation in student exam performance across self-reported deep-strategy scores in an introductory accounting course (Summer 2017 through Spring 2019).

Relative Contribution of Variance Sources

Only two of the six main-effect ANOVA factors significantly accounted for variation in exam scores (see Lines 1, 2, and 6 in Table 1). Deep strategy score accounted for the largest (33.3%; see Line 6, Table 2), with considerably lesser contribution from exam number (see Line 1 in Table 2).

Table 2. Restricted-maximum likelihood estimation of variance components in normalized-rank student exam scores from an introductory accounting course (Summer 2017 through Spring 2019)

Line	Variance Model Term	% of σ^2_{TOTAL}
1	Exam number, EXAM	7.82
2	Academic term, TERM	4.86
3	Student gender, GEND, nested within TERM	0
4	Optional cumulative project student group, PROJ, nested within TERM and GEND	1.93
5	Academic load, LOAD, nested within TERM, GEND, and PROJ	2.84
6	Deep strategy score, DEEP, nested within TERM, GEND, PROJ, and LOAD	33.32
7	EXAM \times TERM interaction	1.41
8	EXAM \times GEND interaction, nested within TERM	0.52

9	EXAM × PROJ interaction, nested within TERM and GEND	0
10	EXAM × LOAD interaction, nested within TERM, GEND, and PROJ	3.98
11	EXAM × DEEP interaction, nested within TERM, GEND, PROJ, and LOAD	0
12	Unexplained (= error) variance	43.32
Total variance, $\sigma^2_{\text{TOTAL}} = 1.070$		100.00

We also found notable non-zero contributions from statistically non-significant variance sources. In descending order, these sources were (a) academic term, (b) interaction between exam number and academic load, (c) academic load, (d) OCP student group, and interactions of exam number with (e) with academic term and (f) student gender.

Discussion

The results do not definitively support our hypothesis that OCP participation predicts subsequent exam scores. Given the general decline in exam scores over the course of a semester, if the results had supported our hypothesis, then exam scores of BMP would have declined at a faster rate than AMP students, while exam scores of OOP students would have declined the fastest. Instead, the lack of interaction between exam number and OCP student group (see Line 9, Table 1), together with the statistical non-significance of OCP student group as a main factor (see Line 4, Table 1), suggests that OCP participation had little influence on the overall trajectory of academic performance across a semester. This particular finding is entirely consistent with Wynn-Williams et al. (2016), who found that experimentally assigned unstructured case studies that they expected to promote deeper learning in accounting students actually had no effect on final course grades.

The slight performance increases we found in Exam 2 and Exam 4 are, therefore, worth additional attention (see Figure 1). We attribute the uptick in Exam 2 scores by AMP students to OCP participation—the small (<2%) variance contribution by the PROJ model term (see Line 4, Table 2) seems to reflect, in part, an immediate but temporary benefit of OCP instruction to AMP students (vs. presumably less engaged students in the other OCP groups). Later in the academic term, after further exposure to discrete elements in the accounting cycle, BMP and OOP students (and AMP students to a much lesser degree) scored higher on Exam 4, perhaps reflecting a better grasp of the accounting cycle. Such nuanced gains, however, were fleeting because scores after Exam 4 declined synchronously across the three OCP student groups (see Figure 1). While attributing post-Exam 4 declines to student stress or distraction from upcoming mid-term exams and project deadlines in other courses is tempting, the lack of significant effects from academic load by itself (see Line 5, Table 1) or from its interaction with exam (see Line 10, Table 1) suggests otherwise. Because many students modify their schedules by dropping and adding courses in the first week of an academic term, perhaps scholars might investigate the effects of academic load by comparing the type and amount of homework per week that might be expected in concurrent math and/or writing-intensive courses.

Perhaps our most surprising result is the significant degree to which self-reported deep strategy scores predicted exam scores (see Line 6, Table 1). While students scoring the highest on the deep-strategy scale tended to perform well across the seven exams, students on the opposite end of the deep strategy scale also performed nearly the same. Notably, one student posting the lowest deep-

strategy score actually performed as well as the eight students scoring 22 and 23 (out of 25) on the deep strategy scale (see Figure 2). A *post hoc* correlation analysis seemed to confirm the contours of a “shallow depression” in Figure 2—that is, a nonlinear relationship between self-reported deep strategy score and Exam 1 score (partial Pearson $r = 0.047$, $p > 0.05$). Together, these ambiguous results seem to agree with the findings of Gijbels et al. (2005) and Choy et al. (2012) regarding the generally weak but positive relationship between deep learning and classroom achievement.

Consistent with our thinking, Turner and Baskerville (2013) suggested that (more) experience of deep learning is a necessary prelude to highly valued social behaviors, including problem solving, teamwork, and communication. Students in the current study—most of whom were in their 2nd year of university study—might have misperceived their own learning intent and study practices, resulting in misleading SAL survey scores. Indeed, Ehrlinger et al. (2008) found that underperforming students overestimated their own abilities, even with substantive monetary incentives for accurate self-assessment. Follow-up studies that measure the relationship between self-reported learning approaches and observed classroom behaviors would be worthwhile.

The current study has some limitations that open pathways to future research. For one, a larger and more robust sample of university students across various course sections taught by different instructors would not only facilitate generalization to other student populations but also allow future researchers to incorporate into their statistical analyses other explanatory variables of academic performance (e.g., prior learning; Jones & Fields, 2001). Larger sample sizes would also allow investigation into explicit student behaviors outside the classroom and demographic factors such as ethnicity, socio-economic status, household income, parental education, job-related work hours, etc. In addition, scholars might consider the degree to which procrastination (Rotenstein et al., 2009) and all-night “cramming” (Hershner & Chervin, 2014) affect exam performance, though earlier findings suggest that cramming benefits achievement- and goal-oriented students (e.g., Brinthaupt & Shin, 2001). Researchers should also investigate student academic performance beyond a single course, what Lorz et al. (2013) referred to as the “stability of the dependent variable” (p. 143). If self-identified “slow learners” learned accounting principles at all, then a longer-term longitudinal study, along the lines of Sargent and Borthick (2013), who analyzed prior learning and student GPA, might more properly document deep learning in the form of improved academic performance in later, upper-division coursework.

Moreover, establishing causality requires actual experimental manipulation, along with proper controls for myriad and possibly confounding factors (see Baeten et al., 2010 for an excellent review). Though perfectly applicable to diverse fields of human sciences, including healthcare and higher education (Thyer, 2012), our quasi-experimental design only allowed us to explore possible variables affecting student academic performance. To the extent allowed by ethical practice, scholars should consider running parallel course sections, whereby only one section would contain the experimental variable of interest (e.g., OCP of the current study).

Finally, without proper replication by researchers in other academic fields, our findings remain (only) discipline- and population-specific. While an optional cumulative project about the accounting cycle explained only 1.93% of variation in exam performance by business majors, similarly immersive and discipline-specific tutorials might have different effects on exam performance and learning outcomes in other fields. For example, as the *raison d'être* of science laboratory courses, immersive simulations and practicums seem indispensable to sound training of future healthcare professionals (Sanko, 2017). Examining board certification exam outcomes as a function of participant presence in training simulations might yield powerful insights (Makransky et al., 2020).

Conclusion

Our findings did not definitively support the hypothesis that voluntary participation by students in an optional cumulative project (OCP) would predict subsequent exam performance. However, we successfully quantified contribution of salient factors, including academic load, OCP participation, and self-reported adoption of deep learning strategies, to student exam performance (i.e., final course grades). Whether the more successful students (with or without the benefit of interactive tutorials about the accounting cycle) actually exercised deep learning strategies that might explain their relatively high exams scores is a question worth addressing in future studies.

Acknowledgements

We thank Steven Freeman and Laura Slane at the Center for the Integration of Research, Teaching, and Learning (based at Iowa State University and University of Wisconsin-Madison, respectively) for motivating the completion of this study. The authors greatly appreciate the very insightful critiques from three anonymous reviewers.

Authorship Statement

The authors declare the following specific contributions: PML (Accounting & Law Department) originated the research study/questions, administered all classroom surveys, collected all student academic data, drafted a working manuscript, and edited subsequent versions of the manuscript. AEO (Biology Department) designed the experiment, performed all statistical analyses and results presentation, drafted a working manuscript, and edited subsequent versions of the manuscript.

References

- Allemann-Ghionda, C. (2000). Dewey in postwar-Italy: The case of re-education. *Studies in Philosophy and Education*, 19, 53–67.
- The Apache Software Foundation. (2014). OpenOffice Calc [computer software]. Retrieved from <https://www.openoffice.org>.
- Baeten, M., Kyndt, E., Struyven, K., and Dochy, F. (2010). Using student-centered learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, 5, 243–260.
- Bati, A.H., Tetik, C., and Gürpınar, E. (2010). Assessment of the validity and reliability of the Turkish adaptation of the study process questionnaire (R-SPQ-2F). *Turkiye Klinikleri Journal of Medical Science*, 30 (5), 1639–1646. [in Turkish, with English-translated abstract]
- Biggs, J.B. (1979). Individual differences in study processes and the quality of learning outcomes. *Higher Education*, 8, 381–394.
- Biggs, J., Kember, D., and Leung, D.Y.P. (2001). The revised two-factor Study Process Questionnaire, R-SPQ-2F. *British Journal of Educational Psychology*, 71, 133–149.
- Brinthaup, T.M. and Shin, C.M. (2001). The relationship of academic cramming to flow experience. *College Student Journal*, 35 (3), 457–472.
- Choy, J.L.F., O’Grady, G., and Rotgans, J.L. (2012). Is the study process questionnaire (SPQ) a good predictor of academic achievement?: Examining the mediating role of achievement-related classroom behaviors. *Instructional Science*, 40, 159–172.

- Christou, T.M. (2013a). "Schools are no longer merely educational institutions": The rhetoric of social efficiency in Ontario education, 1931–1935. *History of Education*, 42 (5), 566–577.
- Christou, T.M. (2013b). The complexity of intellectual currents: Duncan McArthur and Ontario's progressivist curriculum reforms. *Paedagogica Historica*, 49 (5), 677–697.
- Conover, W.J. (1999). *Practical nonparametric statistics* (3rd ed.). New York, NY: John Wiley & Sons.
- Darling, J. and Nisbet, J. (2000). Dewey in Britain. *Studies in Philosophy and Education*, 19, 39–52.
- Duff, A. and McKinstry, S. (2007). Students' approaches to learning. *Issues in Accounting Education*, 22 (2), 183–214.
- Duff, A. and Mladenovic, R. (2015). Antecedents and consequences of accounting students' approaches to learning: A cluster analytic approach. *British Accounting Review*, 47, 321–338.
- Dunkin, M.J. and Biddle, B.J. (1974). *The study of teaching*. New York, NY: Holt, Rinehart, & Winston.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., and Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105, 98–121.
- Entwistle, N. (2015). The origins and evolution of the concept of "approaches to learning." *The Psychology of Education Review*, 39 (2), 9–14.
- Gijbels, D., Van de Watering, G., Dochy, F., and Van den Bossche, P. (2005). The relationship between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education*, 10 (4), 327–341.
- Greenfield, P.M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, 24 (9), 1722–1731.
- Groccia, J.E. (2012). A model for understanding university teaching and learning. In J.E. Groccia, M.A. Alsudairi, and W. Buskist (Eds.), *Handbook of college and university teaching: A global perspective* (pp. 2–13). Thousand Oaks, CA: Sage Publications.
- Guess, A.K. (2014). A methodology for case teaching: Becoming a guide on the side. *Journal of Accounting and Finance*, 14 (6), 113–126.
- Hall, M., Ramsay, A., and Raven, J. (2004). Changing the learning environment to promote deep learning approaches in first-year accounting students. *Accounting Education: An International Journal*, 13 (4), 489–505.
- Healy, M. and McCutcheon, M. (2010). Teaching with case studies: An empirical investigation of accounting lecturers' experiences. *Accounting Education: An International Journal*, 19 (6), 555–567.
- Hershner, S.D. and Chervin, R.D. (2014). Causes and consequences of sleepiness among college students. *Nature and Science of Sleep*, 6, 73–84.
- Immekus, J.C. and Imbrie, P.K. (2010). A test and cross-validation of the revised two-factor study process questionnaire factor structure among western university students. *Educational and Psychological Measurement*, 70 (3), 495–510.
- Jones, J.P. and Fields, K.T. (2001). The role of supplemental instruction in the first accounting course. *Issues in Accounting Education*, 16 (4), 531–547.
- Justicia, F., Pichardo, M.C., Cano, F., Berben, A.B.G., and de la Fuente, J. (2008). The revised two-factor study process questionnaire (R-SPQ-2F): Exploratory and confirmatory factor analyses at item level. *European Journal of Psychology of Education*, 23 (3), 355–372.
- Kolb, A.Y. and Kolb, D.A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4 (2), 193–212.
- Koljatic, M. and Kuh, G.D. (2001). A longitudinal assessment of college student engagement in good practices in undergraduate education. *Higher Education*, 42 (3), 351–371.

- Lake, W., Boyd, Wm., and Boyd, We. (2017). Understanding how students study: The genealogy and conceptual basis of a widely used pedagogical research tool, Biggs' Study Process Questionnaire. *International Education Studies*, 10 (5), 100–108.
- Lorz, M., Mueller, S., and Volery, T. (2013). Entrepreneurship education: A systematic review of the methods in impact studies. *Journal of Enterprising Culture*, 21 (2), 123–151.
- Makransky, G., Andreassen, N.K., Baceviciute, S., and Mayer, R.E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000473>.
- Marton, F. (1976). What does it take to learn? Some implications of an alternative view of learning. In N. Entwistle (Ed.), *Strategies for research and development in higher education* (pp. 32–16). Amsterdam: Swets & Zeitlinger.
- Marton, F. and Säljö, R. (1976a). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46, 4–11.
- Marton, F. and Säljö, R. (1976b). On qualitative differences in learning: II. Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology*, 46, 115–127.
- McKeachie, W.J. (1990) Research on college teaching: The historical background. *Journal of Educational Psychology*, 82 (2), 189–200.
- Michel, N., Cater, J.J., and Varela, O. (2009). Active versus passive teaching styles: An empirical study of student learning outcomes. *Human Resource Development Quarterly*, 20 (4), 397–418.
- Mirghani, H.M., Ezimokhai, M., Shaban, S., and van Berkel, H.J.M. (2014). Superficial and deep learning approaches among medical students in an interdisciplinary integrated curriculum. *Education for Health*, 27 (1), 10–14.
- Nelson Laird, T.F., Seifert, T.A., Pascarella, E.T., Mayhew, M.J., and Blaich, C.F. (2014). Deeply affecting first-year students' thinking: Deep approaches to learning and three dimensions of cognitive development. *Journal of Higher Education*, 85 (3), 402–432.
- Neter, J., Wasserman, W., and Kutner, M.H. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Richard D. Irwin.
- Olvido, A.E. and Wagner, W.E., Jr. (2004). Signal components, acoustic preference functions and sexual selection in a cricket. *Biological Journal of the Linnean Society*, 83, 461–472.
- Olvido, A.E., Fernandes, P.R., and Mousseau, T.A. (2010). Relative effects of juvenile and adult environmental factors on mate attraction and recognition in the cricket, *Allonemobius socius*. *Journal of Insect Science*, 10 (90). <https://doi.org/10.1673/031.010.9001>.
- Pelaez, N., Anderson, T.R., and Postlethwait, S.N. (2015). A vision for change in bioscience education: Building on knowledge from the past. *BioScience*, 65 (1), 90–100.
- Phillips, F. and Heiser, L. (2011). A field experiment examining the effects of accounting equation emphasis and transaction scope on students learning to journalize. *Issues in Accounting Education*, 26 (4), 681–699.
- Richardson, J.T.E. (2015). Approaches to learning or levels of processing: What did Marton and Säljö (1976a) really say? The legacy of the work of the Göteborg group in the 1970s. *Interchange*, 46, 239–269.
- Roith, C. (2014). Educational theory and practice in post-revolutionary times: The European academic debate on the experimental schools in Hamburg (1919–1933) in the 1930s and 1970s. *Paedagogica Historica*, 50 (5), 631–650.
- Rotenstein, A., Davis, H.Z., and Tatum, L. (2009). Early birds versus just-in-timers: The effect of procrastination on academic performance of accounting students. *Journal of Accounting Education*, 27 (4), 223–232.

- Sanko, J.S. (2017). Simulation as a teaching technology: A brief history of its use in Nursing education. *Quarterly Review of Distance Education*, 18 (2), 77–85.
- Sargent, C.S. and Borthick, A.F. (2013). Evidence for insisting on cognitive conflict tasks: Impact on accounting majors in upper-level courses. *Issues in Accounting Education*, 28 (4), 759–777.
- SAS Institute. (1999). *Selected SAS® documentation for biometry 902: Advanced experimental design* (2nd ed.). Cary, NC: SAS Institute.
- Saudagaran, S.M. (1996). The first course in accounting: An innovative approach. *Issues in Accounting Education*, 11 (1), 83–94.
- Schneider, J.H. (2000). John Dewey in France. *Studies in Philosophy and Education*, 19, 69–82.
- Slavich, G.M., and Zimbardo, P.G. (2012). Transformational teaching: Theoretical underpinnings, basic principles, and core methods. *Educational Psychology Review*, 24 (4), 569–608.
- Sokal, R.R. and Rohlf, F.J. (1981). *Biometry* (2nd ed.). San Francisco, CA: Freeman and Associates.
- Soper, D.S. (2018). *p-Value Calculator for an F-Test* [Software]. Retrieved from <https://www.danielsoper.com/statcalc>.
- Stes, A., De Maeyer, S., and Van Petegem, P. (2013). Examining the cross-cultural sensitivity of the revised two-factor study process questionnaire (R-SPQ-2F) and validation of a Dutch version. *Public Library of Science ONE*, 8 (1), e54099. <https://doi.org/10.1371/journal.pone.0054099>.
- Stolk, V., Los, W., and Karsten, S. (2014). Education as cultural mobilisation: The Great War and its effects on moral education in the Netherlands. *Paedagogica Historica*, 50 (5), 685–706.
- Svensson, A.M. (1976). *Study skills and learning*. Goteberg, Sweden: Acta Universitates, Gothenburgensis.
- Thyer, B.A. (2012). *Quasi-experimental research designs*. New York, NY: Oxford University Press.
- Turner, M. and Baskerville, R. (2013). The experience of deep learning by accounting students. *Accounting Education: An International Journal*, 22 (6), 582–604.
- U.S. Department of Education, National Center for Education Statistics, Higher Education General Information Survey (HEGIS). (2013). “Degrees and Other Formal Awards Conferred” surveys, 1970–71 through 1985–86; Integrated Postsecondary Education Data System (IPEDS), “Completions Survey” (IPEDS-C: 91–96); and IPEDS Fall 2001 through Fall 2012, Completions component. (Table 318.20 was prepared July 2013).
- Warren, C.S., Reed, J.M., and Duchac, J. (2014). *Financial & managerial accounting* (12th ed.). Mason, OH: South-Western Cengage Learning.
- Warren, D.L. and Young, M.N. (2012). Integrated accounting principles: A best practices course for introductory accounting. *Issues in Accounting Education*, 27 (1), 247–266.
- Wawrzynski, M. and Baldwin, R. (2014). Promoting high-impact student learning: Connecting key components of the collegiate experience. *New Directions for Higher Education*, 2014 (165), 51–62.
- Wynn-Williams, K., Beatson, N., and Anderson, C. (2016). The impact of unstructured case studies on surface learners: A study of second-year accounting students. *Accounting Education*, 25 (3), 272–286.
- Zakariya, Y.F., Bjørkestøl, K., Nilsen, H.K., Goodchild, S., and Loras, M. (2020). University students’ learning approaches: An adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Educational Evaluation*, 64 (2020), 100816. <https://doi.org/10.1016/j.stueduc.2019.100816>.