RETURNS TO TEACHING REPETITION: THE EFFECT OF SHORT-TERM TEACHING EXPERIENCE ON STUDENT OUTCOMES

Harold E. Cuffe

School of Economics and
    Finance
Victoria University of
    Wellington
Wellington 6140, New Zealand
harold.cuffe@vuw.ac.nz

Jan Feld

(corresponding author)
School of Economics and
    Finance
Victoria University of
    Wellington
Wellington 6140, New Zealand
jan.feld@vuw.ac.nz

Trevor O'Grady

Department of Economics
School of Business
The College of New Jersey
Ewing, NJ 08628
ogradyt@tcnj.edu

Abstract

Teachers often deliver the same lesson multiple times in one day. In contrast to year-to-year teaching experience, it is unclear how this teaching repetition affects student outcomes. We examine the effects of teaching repetition in a higher education setting where students are randomly assigned to a university instructor's first, second, third, or fourth lesson on the same day. We find no meaningful effects of repetition on grades, course dropout, or study effort and only suggestive evidence of positive effects on teaching evaluations. These results suggest that teaching repetition is a powerful tool to reduce teachers' preparation time without negative effects on students.

1. INTRODUCTION

Assigning a single teacher to teach multiple sections of a course is a common practice meant to reduce the costs of delivering course content. However, the consequences for students of this time-saving arrangement—a practice we refer to as teaching repetition—are not well understood.

It is not immediately obvious whether teaching repetition benefits or harms students. One possibility is that teachers simply warm up in the first section, and deliver the material more fluently in subsequent repetitions. Teaching repetition also allows teachers to learn on the job. For example, teachers who repeat the same lesson might incorporate student feedback from earlier sections. There is abundant evidence that year-to-year teaching experience positively affects teaching effectiveness (for review, see Harris and Sass 2011). We may therefore find evidence that these persistent improvements occur swiftly after each repetition. However, teaching repetition may also lead to worse student outcomes. The monotony of teaching the same lesson multiple times may lead to mental fatigue and a worse learning experience for the student. A lack of variety in teaching or other tasks is thought to be an important contributor to instructor "burnout" (see, for example, Kaikai and Kaikai 1990 or Maslach, Schaufeli, and Leiter 2001). More generally, neuroscientific evidence suggests that task repetition–related mental fatigue adversely affects performance, motivation, and error correction in that task (Lorist, Boksem, and Ridderinkhof 2005; Boksem, Meijman, and Lorist 2006).

Beyond academic research, there is wider recognition that teaching repetition may entail tradeoffs related to teaching quality. For example, the Association of Departments of English recommends: "In general, the proper number of *different* courses likely to ensure excellent teaching is two or three; that is, there should be enough variety to promote freshness but not so much as to prevent thorough preparation"[1] (emphasis added). This recommendation highlights the perceived tension between time savings and the tedium of teaching repetition.

In this paper, we use a large administrative dataset from a Dutch business school to test how teaching repetition affects students' grades, dropout rates, how students evaluate their instructors, and the amount of time they put into studying for the course. Our empirical analysis focuses on comparisons of student outcomes across an instructor's sections of a course for a given term. Importantly, students in our setting are randomly assigned to sections within each course for which they are registered. After accounting for possible confounding variables, such as a section's start time, we interpret any observed differences in average outcomes for students in an instructor's later sections as the causal effects of teaching repetition.

Overall, our results show little evidence that teaching repetition benefits or harms students. In our main specification, most of our point estimates are small and none is statistically significant. For students in an instructor's second section (relative to the first section), we can rule out effects on grades that are below −5.1 percent and above 3.4 percent of a standard deviation, based on the 95 percent confidence interval. In specifications where we do not control for section starting time (which is highly collinear with the number of repetitions), our effects are similar and even more precisely

---

1. See https://www.ade.mla.org/Resources/Policy-Statements/ADE-Guidelines-for-Class-Size-and-Workload-for-College-and-University-Teachers-of-English-A-Statement-of-Policy.

estimated. Here we can rule out economically meaningful effects on grades as well as student dropout rates and study hours. From the 95 percent confidence intervals, we can rule out effects larger than 3 percent of a standard deviation on grades, 1 percentage point on the probability of dropping out of the course, and 33 minutes of further self-study per week. In these specifications, however, we do find that teaching repetition improves teaching evaluations between 3.4 and 5.3 percent of a standard deviation. While not affecting students' objective academic outcomes, teaching repetition may allow instructors to deliver the material in a way that is appreciated by students. We also find suggestive evidence that the positive impact of repetition on teaching evaluations is larger for inexperienced instructors. Finally, we see no evidence that the effect of teaching repetition is different if the instructor had a break before having to teach a subsequent section. These results suggest that any adjustments to the course material instructors make in such a break does not affect their teaching effectiveness, and that short-term instructor fatigue is not a significant determinant of student outcomes.

We are one of the first studies to empirically examine the consequences of teaching repetition. The only other study is by Williams and Shapiro (2018), who use data from the United States Air Force Academy to investigate how student fatigue, time of instruction, and teaching repetition affect student outcomes. For identification, they also rely on random assignment of students to sections within a course. Their results show small positive effects of teaching repetition: Students who are in an instructor's second compared to first section achieve 3 percent of a standard deviation higher grades. Whereas Williams and Shapiro examine how multiple aspects of the university schedule affect students' grades, we focus on the effect of teaching repetition. In our more thorough analysis of the teaching repetition aspect of scheduling, we consider a number of important outcomes beyond student grades and investigate heterogenous effects along a number of dimensions. The results of Williams and Shapiro and our study together allow us to draw robust conclusions: Teaching repetition does not harm students and has, if anything, only small positive effects. Universities can continue to benefit from the efficient use of staff time with schedules that allow for teaching repetition.

Besides the paper's direct policy relevance, our findings help reveal how teaching experience affects teacher productivity. Teaching repetition can be viewed as an intensive way of accumulating curriculum-specific experience, which has been shown to improve teaching effectiveness. For example, Ost (2014) uses data on fifth-grade teachers in North Carolina to show that curriculum-specific experience improves a teacher's effectiveness, particularly for mathematics, even after controlling for general teaching experience. At the postsecondary-level, De Vlieger, Jacob, and Stange (2018) use data from the University of Phoenix, a for-profit university, and find that teaching effectiveness in college algebra is positively related to curriculum-specific experience (sections taught) but uncorrelated with length of tenure at the university. In contrast to our study, these studies do not distinguish whether a teacher taught the same subject over a long period of time or within the same day. Our findings are consistent with work on the psychology of learning that shows that "re-studying a piece of information immediately after the first study episode is not an efficient way to proceed in order to learn effectively" (Gerbier and Toppino 2015, p. 50). Similarly, teaching productivity does not improve merely from quickly repeated delivery of course content, and instead must occur over a time horizon that allows for substantive reflection and reaction.

2. BACKGROUND

Empirical Setting

Our data come from a Dutch business school and cover the academic years 2009–10 to 2014–15.[2] This business school offers bachelor's, master's, and PhD programs in business studies, economics, finance, and econometrics. An academic year consists of four regular teaching terms of eight weeks each. In these terms, students typically take two courses at the same time. For brevity, we refer to each course-term-year combination as a course. For example, we refer to Economics 101 taught in term 1 of 2011 and Economics 101 taught in term 1 of 2012 as separate courses. In a typical course, all students attend three to seven lectures together and twelve two-hour tutorial meetings in sections of up to sixteen students. In this paper, we focus on the effects of repeated teaching across sections of these tutorials over the course of one day.

We exclude a number of observations due to deviations from the standard scheduling procedure, apparent mistakes in the data, or missing data on key control variables.[3] After these exclusions, we observe 10,898 students and 83,195 student course enrollments.

Table 1 panel A shows summary statistics for our estimation sample. Thirty-nine percent of students are female. The average age of students is 21.2 years and a majority of them are either Dutch (31 percent) or German (44 percent). On average, we observe each student for 7.6 different courses in our estimation sample.

We also observe 731 different instructors, who vary in their seniority from bachelor's and master's students (46 percent), and PhD students (23 percent), to more senior instructors including postdocs, lecturers, and assistant, associate, and full professors (31 percent). Each instructor teaches between one and four sections, with the average teaching load being 2.5 sections per course. In more than 99 percent of the cases, instructors teach all of their sections within a single day.

All sections in a given course cover the same material and have the same assignments. For a typical section meeting, students discuss with their section peers assigned readings or solutions to exercises. Students are expected to prepare the course material beforehand. Instructors are expected to prepare the same material thoroughly enough that they are able to answer students' questions and to structure the session by, for example, deciding the order in which to discuss the course material. The main role of the tutorial instructor during the section meetings is to guide the discussion and help students when they are stuck.

We observe 7,292 total sections. Figure 1 shows that of these 7,292 sections, 42 percent are an instructors' first section for a given course (2,900). Thirty-three percent (2,431), 19 percent (1,411), and 8 percent (550) are an instructor's second, third, and fourth sections, respectively. This is the variation we exploit to estimate the effect of teaching repetition. We draw from this figure that there exists a nontrivial number of students receiving course material as part of an instructor's third or fourth repetition. However, the unequal distribution of observations across repetitions suggests we should expect more precise estimates of the effects of one or two repetitions relative to three or four.

---

2. For more detailed information on the institutional environment, see Feld and Zölitz (2017) and Feld, Salamanca, and Zölitz (2020).

3. For more details see the online appendix, which can be accessed on *Education Finance and Policy*'s Web site at https://doi.org/10.1162/edfp_a_00309.

Table 1. Descriptive Statistics

|	N (1)	Mean (2)	SD (3)	Minimum (4)	Maximum (5)	ρ (6)
Panel A: Individual Characteristics						
Student level						
Female	10,898	0.39	0.49	0.00	1.00	
Age, years	10,898	21.17	2.50	15.93	44.25	
Dutch	10,898	0.31	0.46	0.00	1.00	
German	10,898	0.44	0.50	0.00	1.00	
Bachelor's student	10,898	0.64	0.44	0.00	1.00	
Courses per student	10,898	7.63	6.27	1.00	33.00	
Instructor level						
Student instructor	731	0.46	0.49	0	1	
PhD student instructor	731	0.23	0.41	0	1	
Senior instructor	731	0.31	0.45	0	1	
Instructor-course level						
Sections per course	2,928	2.49	0.98	1.00	4.00	
Panel B: Student Outcomes						
Academic outcomes						
Grade	77,269	6.70	1.76	1	10	
Dropout	83,195	0.07	0.26	0	1	
Course evaluation survey responses						
Evaluate the overall functioning of your tutor in this course with a grade (1–10)	27,144	7.77	1.98	1	10	0.94
The tutor sufficiently mastered the course content (1–5)	27,144	4.31	0.95	1	5	0.82
The tutor stimulated the transfer of what I learned in this course to other contexts (1–5)	27,144	3.94	1.08	1	5	0.87
The tutor encouraged all students to participate in the (tutorial) group discussions (1–5)	27,144	3.60	1.18	1	5	0.74
The tutor was enthusiastic in guiding our group (1–5)	27,144	4.07	1.10	1	5	0.87
The tutor initiated evaluation of the group functioning (1–5)	27,144	3.64	1.22	1	5	0.68
Self-study hours per week	26,918	14.25	8.32	0	90	

Notes: This table is based on our estimation sample. Column 6 reports the correlation between our composite index for course evalutions with each of its six consitutent items. SD = standard deviation.

Outcome Variables

We investigate four outcomes related to teaching effectiveness or perceived teaching effectiveness: (1) a student's grade in the course, (2) an indicator for whether the student dropped out of the course, (3) an index of student evaluations of the instructor across several dimensions, and (4) a student's self-reported study hours per week for the course. Table 1 panel B shows summary statistics for these outcomes.

Course grades often consist of multiple graded components, such as the presentation grade, participation grade, or final exam grade. The graded components and their weights differ by course, with most weight usually given to the final exam. In a typical course, final exams are graded by the course coordinator and all section instructors, with each grading the same set of exam questions for all students in the course. Student participation and presentations are typically graded by their section instructor, but this usually constitutes a small part of students' overall grades.

Course grades are assigned on a scale from 1 to 10, with 5.5 being the lowest passing grade. The average grade in our sample is 6.7. To facilitate the interpretation of results
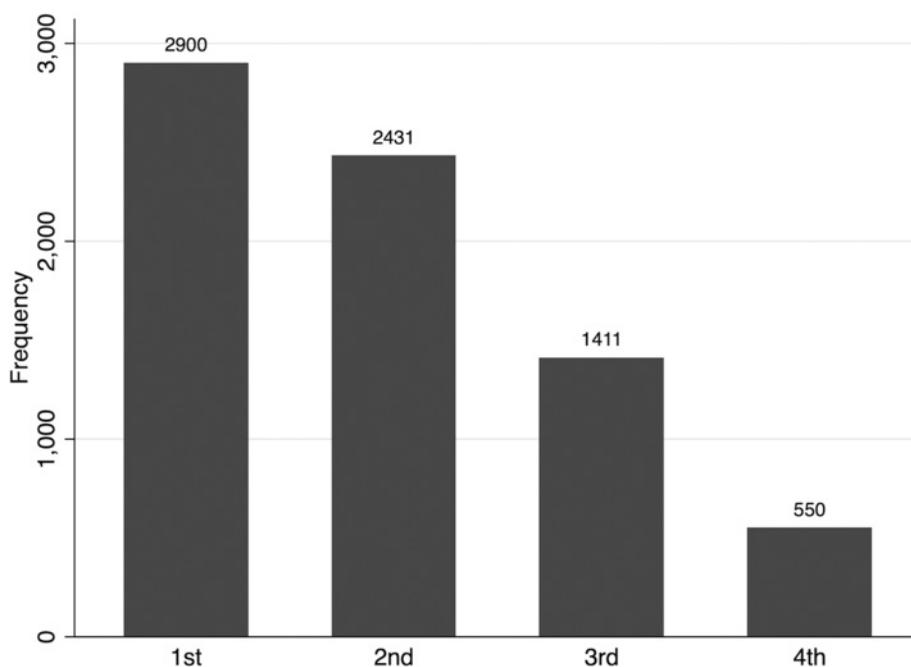
Note: Figure 1 shows the number of sections in our estimation sample that are an instructor's first, second, third, and fourth sections of a given course.

Figure 1. Number of Sections of Different Order

in our empirical analysis, we standardize course grades to have a mean of zero and standard deviation of 1 over the estimation sample.

Students drop out when they register for a course but their final grades are missing in the official records. The dropout rate for our sample is 7 percent.

Students are prompted to fill out course evaluations at the end of the term, which include questions about the course, the instructor, and the student's experiences in the class. Generally, teaching evaluations gauge students' satisfaction with instructors and courses but are not a direct measure of teaching effectiveness. Indeed, there is ample evidence that comparing teaching evaluations across instructors is a poor measure of their relative effectiveness as a teacher (Uttl, White, and Gonzalez 2017). Despite this, changes in an instructor's teaching evaluations across section repetitions can reveal qualities of the classroom experience that evolve as an instructor repeats lesson material (e.g., how instructors gain apparent confidence with the material or lose enthusiasm from fatigue), without reflecting fixed characteristics of the instructor (e.g., grading style, attractiveness). For universities and instructors, there is additional cause for understanding the determinants of student evaluation scores because promotion and retention are often tied to such measures.

We use six questions to measure the instructors' teaching effectiveness. These questions measure instructors' (1) overall functioning, (2) mastery of the course content, (3) ability to transfer course content to other contexts, (4) encouragement of student participation, (5) enthusiasm in guiding the group, and (6) whether the instructors initiated the evaluation of the group functioning (see panel B of table 1 for the wording of

the instructor evaluation items). Table A.1 in the online appendix gives the correlation matrix for these six variables. All of the variables positively correlate with one another, with the strongest correlation occurring between instructors' overall functioning and mastery of the course content (questions 1 and 2), and the weakest correlation between instructors' mastery of the course content and their initiation of the evaluation of the group functioning (questions 2 and 6).

To broadly assess how teaching repetition affects students' perceptions of an instructor's effectiveness, we first combine these evaluation variables using principal factor analysis. This exercise identifies a single principal factor, which we standardize to have mean zero and a standard deviation of 1, and use as our dependent variable measuring student's subjective assessment of instructor performance.

To measure self-study hours, we use the students' answers to the question of how many hours they studied (excluding time in lectures and tutorials).

Throughout the empirical analysis we use the maximum sample size possible for each student outcome. The sample for the dropout indicator includes everyone initially enrolled in the course, while the sample for course grades only includes those completing the course (93 percent of enrollees). Because responding to course evaluations is voluntary, the sample of instructor evaluation scores and study hours only includes students who chose to answer these questions on the course evaluation surveys at the end of the term (33 percent and 32 percent of enrolled students, respectively). Table A.2 in the online appendix shows that female students and students with higher grade point averages (GPAs) are more likely to respond to course evaluations, as well as some heterogeneity in response by nationality. This selective response implies that our effect estimates for these latter outcomes may not be representative of the broader student population. Importantly, however, section-order does not predict responses to instructor evaluations and study hours questions.

Assignment of Instructors and Students to Sections

An advantage of our setting is that students are randomly assigned to sections within a course conditional on scheduling conflicts. Scheduling conflict arises for about 5 percent of student-course registrations and are resolved by schedulers manually switching students between sections. From the academic year 2010–11, the business school additionally stratifies section assignment in bachelor's courses by student nationality to encourage a mixing of Dutch students and German students. Other papers using this dataset have shown that student assignment to sections has the properties we would expect under random assignment (e.g., Feld, Salamanca, and Zölitz 2020). Instructors are assigned by schedulers to different sections within a course. For this assignment, schedulers do not consider the characteristics of the students in the sections. About 10 percent of instructors indicate a time during which they are not available for teaching. While these constraints potentially affect instructors' time slots, the conditionally random assignment of students to sections ensures that students' characteristics will not predict whether they are in an instructor's first, second, third or fourth section.

Table 2 reports estimates from a regression of students' pre-enrollment characteristics on section order indicators, and controls for section start time and instructor-course-parallel-course fixed effects. Out of the twelve coefficients estimated, we see no statistically significant differences in characteristics at the 5 percent level and only one

Table 2. Randomization Check

Dependent Variable	Grade Point Average (1)	Female (2)	ID Rank (3)	Age (4)
2nd section	−0.045 (−0.111 to 0.021)	0.012 (−0.011 to 0.035)	−117.934 (−331.298 to 95.430)	0.021 (−0.064 to 0.106)
3rd section	−0.092 (−0.208 to 0.024)	0.014 (−0.028 to 0.055)	−280.668 (−647.187 to 85.851)	0.099 (−0.044 to 0.242)
4th section	−0.138* (−0.302 to 0.026)	0.035 (−0.024 to 0.095)	−321.486 (−832.994 to 190.022)	0.126 (−0.076 to 0.327)
Observations	83,195	83,195	83,195	83,195
R^2	0.213	0.176	0.147	0.558
Section 1 average outcome	6.627	.37	7,062.193	20.935
p-value joint significance of all section variables	.4294	.3154	.4306	.1305

Notes: All regressions include instructor-course-parallel-course fixed effects and indicator variables for section starting times. 95 percent confidence intervals based on standard errors clustered at the course-level are in parentheses.

*$p < 0.1$.

at the 10 percent level. While student GPA is marginally lower in the fourth section relative to the first, the section-order coefficients from the GPA regression (or any other regression in table 2) are not jointly significant. Overall, these results show that pre-enrollment characteristics across an instructor's sections are roughly balanced, conditional on controls.

3. EMPIRICAL METHODOLOGY

A number of challenges arise when estimating the causal effects of teaching repetition on student outcomes. For one, instructors teaching multiple sections of a course may be systematically different from instructors who do not. For instance, a more senior and experienced instructor may have a smaller course load and fewer repetitions than an inexperienced instructor. Similarly, teaching repetition may be more common in certain subject areas than others. Because instructor type and course subject are both likely to impact our student outcomes of interest, our analysis only compares student outcomes within instructor-course combinations.[4]

At institutions where students are in full control of their schedule, we may also be concerned about self-selection into earlier or later sections. This problem is largely alleviated by our empirical setting in which students are randomly assigned to sections within a course absent any scheduling constraints. Such scheduling constraints, however, may introduce bias in our estimates. For example, students taking a particularly difficult parallel course that is only offered in the morning may be more likely to end up in an instructor's later section. These students may be relatively high-achievers compared with their peers in earlier sections taking easier parallel courses (introducing positive bias), or these students may have higher workloads and less time to study (introducing negative bias). To account for this potential bias caused by scheduling

---

4. Because the same subject taught in different terms is classified as separate courses in our data, our approach also precludes making comparisons across terms.

constraints, we further restrict comparisons to be between students registered for the same parallel course.

We implement this strategy by estimating regression equation:

$$y_{ijcsd} = \sum_{\tau=2}^{4} \beta_\tau section_{\tau jcs} + \lambda_{jcd} + \delta W_{jcs} + \gamma Z_{ic} + \varepsilon_{ijcsd}, \tag{1}$$

in which y_{ijcsd} is student i's outcome for instructor j's sth section of course c for the term. The subscript d indicates that the student is also registered for parallel course d in that term.

The variable $section_{\tau jcs}$ is a binary indicator for section $\tau > 1$ that takes the value of 1 when $s = \tau$ and zero otherwise.[5] The β_τ parameters represent teaching repetition effects as they measure the difference in outcomes in the τth section relative to the first section for a given instructor-course combination.

The term λ_{jcd} is an instructor-by-course-by-parallel course fixed effect. Our identification of teaching repetition effects therefore relies only on comparisons between students in an instructor's later sections and their peers in the first section who have the same course plan as them. This flexible approach not only accounts for potential sources of bias discussed above, but also any interactions among those sources. We will also show that our results are similar when we only include instructor-by-course fixed effects.

One additional identification concern is that section order is correlated with tutorials' start times. For instance, studies such as that of Williams and Shapiro (2018) find that students tend to perform worse earlier in the day. As section repetitions necessarily come later in the day than the first section, we may mistake these time-of-day effects for repetition effects. Therefore we also control for W_{jcs}, a vector of indicator variables for what time-of-day the section meets.

The vector Z_{ic} consists of student characteristics. These include indicator variables for each student's gender and nationality and cubic polynomials for students' GPA and age at the start of the course. Lastly, ε_{ijscd} is a mean zero error term. In all regressions, we estimate robust standard errors adjusted for clustering at the course level.

4. RESULTS

Main Results

We begin by estimating the effects of teaching repetition on standardized grades. The estimates in column 1 of table 3 support the conclusion that instructors cannot take experience gained in one section, and quickly apply it in subsequent sections in a way that improves student performance. Average grades in instructors' second sections are actually lower than in their first, but the decrease is less than 1 percent of a standard deviation and not statistically significant. The 95 percent confidence interval of this estimate allows us to rule out effects below −5.1 percent and above 3.4 of a standard deviation. For comparison, Williams and Shapiro (2018) find a 3 percent of a standard deviation improvement in average student grades for instructors' second section

---

5. For example, if the student is in section $s = 2$, this means the student is taught by an instructor who has already run through the material once that day. In this case, the entire summation reduces to β_2.

Table 3. The Effects of Teaching Repetition on Student Outcomes

Dependent Variable	Standard Grade (1)	Dropout (2)	Standard Evaluation (3)	Hours (4)
2nd section	−0.008	0.004	0.050	−0.302
	(−0.051 to 0.034)	(−0.008 to 0.016)	(−0.065 to 0.166)	(−1.123 to 0.519)
3rd section	0.012	0.008	0.083	−0.839
	(−0.064 to 0.087)	(−0.014 to 0.030)	(−0.118 to 0.284)	(−2.226 to 0.548)
4th section	0.018	0.009	0.132	−1.159
	(−0.088 to 0.124)	(−0.021 to 0.040)	(−0.142 to 0.407)	(−3.135 to 0.816)
Observations	77,269	83,195	27,144	26,918
R^2	0.569	0.290	0.551	0.412
Section 1 average outcome	.028	.072	−.033	14.415
p-value joint significance of all section variables	.4221	.9020	.7966	.5536

Notes: All regressions include fixed effects for instructor-course-parallel-course combinations and section starting time, cubic polynomials for student grade point average and age, and indicators for student gender and nationality. 95 percent confidence intervals based on standard errors clustered at the course level are in parentheses.

compared with their first.[6] Similarly, De Vlieger, Jacob, and Stange (2018) find that students perform 3 to 4 percent of a standard deviation better on the final exam if the instructor has taught the course at least once before.

Point estimates for an instructors' third and fourth sections relative to their first suggest positive impacts of repetition on grades, but effect sizes are small and are not statistically significant at conventional levels. The confidence intervals for these subsequent repetitions are also considerably larger, which is unsurprising given fewer observed instances of instructors teaching three and four sections of a course. We fail to reject the null hypothesis of a joint test of significance that all section indicator variables equal zero (reported in the final row of table 3).

Although we do not find evidence that teaching repetition affects student grades in the course, teaching repetition may still affect students in other ways. For example, instructors who are better at maintaining student interest might see fewer students drop out midway through the term and receive higher teaching evaluations from their students. In column 2 of table 3, we report estimates of teaching repetition effects for a linear probability model of dropout. Overall, we find small and statistically insignificant effects of teaching repetition on the probability of course dropout. Comparing second to first sections, we can rule out effects on dropout rate below −0.8 and above 1.6 percentage points. Point estimates continue to be small for other sections (less than 1 percentage point) but are measured with slightly less precision.

Similarly, in column 3, we find little evidence that teaching repetition leads to better teaching evaluations. Even as the point estimates rise slightly with repetition, the *p*-value for the joint significance of our section order variables indicates the absence of a strong systematic relationship.

---

6. Our methodology is similar to Williams and Shaprio but not identical. We also estimated additional specifications that more closely align with theirs (unreported), and did not find evidence that our methodological differences drive the differences in our results. These additions included adopting the authors' assumptions regarding instructor fixed effects and clustering.

It remains possible that we do not observe effects of teaching repetition because of students' offsetting behavior. This might occur, for example, if first-section students increase their independent study time to compensate for poorer instructional quality. In column 4, we consider teaching repetition's effects on self-reported weekly study hours. The estimated effects on study hours are small and statistically insignificant, with second-section students spending only approximately 2 percent less time (18 minutes) studying each week, relative to first section students. The point estimates rise somewhat, as do their standard errors, for subsequent sections. However, as before, the effect sizes are small, and the section order variables are not jointly significant. This indicates that students across all sections devote similar amounts of time to study.

Robustness

We probe the robustness of our main results with two additional specifications. First, we estimate the model reported in table 3 without controls for section starting time, which are highly collinear with teaching repetitions. Second, we relax our sample restrictions and include fewer control variables. More specifically, in this second specification, we only exclude observations that represent an exception to the standard section assignment procedure at the business school and observations where the instructor teaches more than four sections in a given course. This leaves us with a substantially larger estimation sample of 107,661 student-course observations (see the online appendix for the sample restrictions). In this specification, we only control for instructor-course fixed effects, effectively comparing mean outcomes of students in the same course taught by the same instructor.

Figure 2 compares the point estimates from the two modified specifications described above to the baseline estimates from table 3 (for point estimate values, see table A.3 in the online appendix). The effects in the alternative specifications are much more precisely estimated than the baseline effects, as evidenced by narrower confidence intervals. The results provide additional support that teaching repetition has no economically significant effect on students' grades, dropout rates, and study effort.

We do, however, see positive and statistically significant repetition effect estimates on teaching evaluations in these specifications. This finding occurs because of increased precision, not because of increases in point estimates. The estimated effects of being in an instructor's second, third, and fourth sections on instructor evaluations are between 3.4 and 5.3 percent of a standard deviation. When estimating the effect of teaching repetition on each evaluation item separately, we show that the positive point estimates are driven by instructors receiving better scores on overall evaluation, content mastery, and ability to transfer what students learned to other contexts (see figure A.1 in the online appendix). Although these estimates may be influenced by section starting time, we interpret them as suggestive evidence that teaching repetition leads to more positive teaching evaluations.

One concern for the interpretation of our results is that section-level curving of presentations and participation may attenuate the effect of teaching repetition on student grades. For these graded components, the instructor may intentionally adjust grades to ensure similar averages across all sections they teach. If student grades on
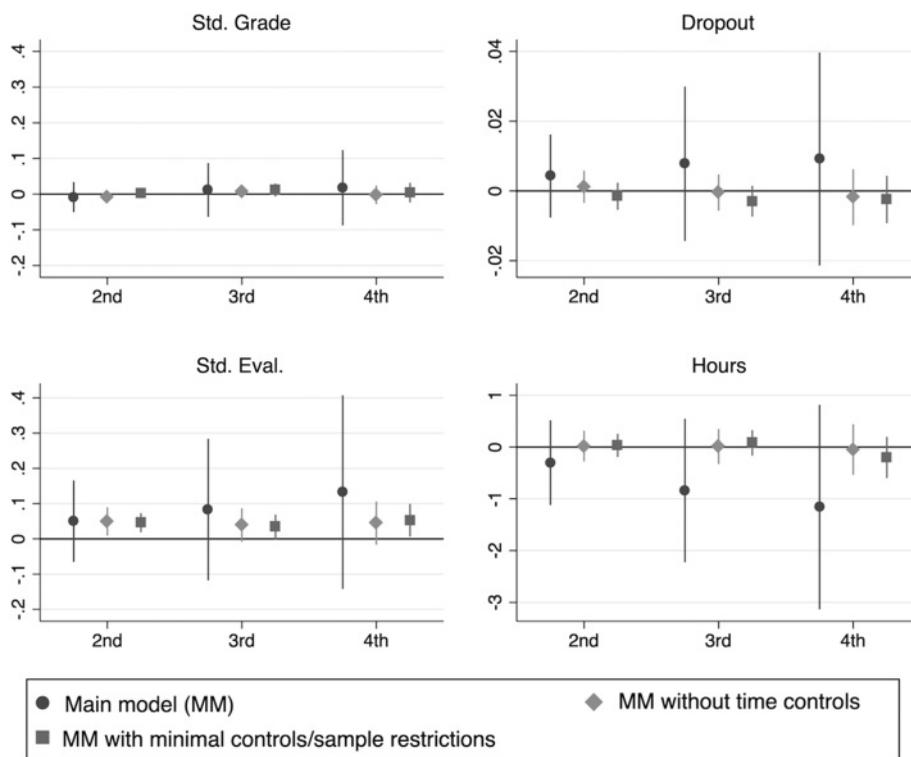
Notes: This figure shows estimates from three specifications for each dependent variable. The point estimates from the first specification are from our main results reported in table 3. The point estimates from the second specification are from regressions without controls for section starting time. The point estimates from the third specification are from regressions with minimal sample restrictions and no controls except for instructor-course fixed effects. Vertical lines indicate 95 percent confidence intervals based on standard errors clustered at the course level.

Figure 2. Robustness

presentations and participation are affected by teaching repetition, section-level curving would obscure this part of the effect on course grades.[7]

To address this concern, we separately estimate the effect of teaching repetition on grades in first-year courses in which grades are entirely based on final exam performance and therefore unaffected by curving at the section level. Results in panel A of online table A.4 show slightly larger effect size estimates for this sample (though still small; 1 percent to 8 percent of a standard deviation) and not statistically significant (*p*-value of joint test: 0.36) suggesting that the absence of effects on grades in our main model is not driven by section-level curving.[8]

---

7. Curving at the course-level may also affect the size of the estimated repetition effect on standardized grades, specifically if the curving method is not a simple linear transformation of raw scores. For example, if the grades for failing students are increased to just above the failing threshold, this would result in a compression of the observed grade distribution even after standardization, which could lead to attenuated repetition effect measurements relative to the effects present in the raw scores. Therefore, our results should be interpreted as the effect of teaching repetition on observed grades that may or may not be curved.

8. For completeness, online table A.4 also shows results for our other outcomes as well as a sample of only non-first-year courses. We find no evidence of a statistically significant repetition effect (based on joint tests) in any of these models.

Table 4. Heterogeneous Effects by Instructor Academic Rank

Dependent Variable	Standard Grade (1)	Dropout (2)	Standard Evaluation (3)	Hours (4)
Panel A: Student and PhD Student Instructors				
2nd section	−0.027 (−0.081 to 0.027)	0.005 (−0.010 to 0.020)	0.155** (0.028 to 0.283)	0.606 (−0.518 to 1.731)
3rd section	−0.022 (−0.115 to 0.070)	0.010 (−0.019 to 0.038)	0.238** (0.043 to 0.433)	0.450 (−1.363 to 2.263)
4th section	−0.030 (−0.161 to 0.101)	0.019 (−0.019 to 0.057)	0.294** (0.018 to 0.570)	0.549 (−1.986 to 3.084)
Observations	38,678	41,916	13,228	12,983
R^2	0.579	0.283	0.550	0.386
Section 1 average outcome	−.044	.078	−.166	13.88
p-value joint significance of all section variables	.5735	.6913	.1026	.4563
Panel B: Senior Instructors				
2nd section	0.021 (−0.049 to 0.091)	0.003 (−0.015 to 0.021)	−0.043 (−0.259 to 0.173)	−1.086 (−2.443 to 0.272)
3rd section	0.066 (−0.058 to 0.190)	0.004 (−0.029 to 0.037)	−0.044 (−0.457 to 0.369)	−1.786 (−4.193 to 0.620)
4th section	0.095 (−0.076 to 0.266)	−0.004 (−0.050 to 0.042)	0.026 (−0.550 to 0.601)	−2.412 (−5.861 to 1.036)
Observations	38,591	41,279	13,916	13,935
R^2	0.559	0.300	0.547	0.437
Section 1 average outcome	.096	.066	.085	14.883
p-value joint significance of all section variables	.5242	.5973	.4040	.4766

Notes: All regressions include instructor-course-parallel-course fixed effects. Additional controls include cubic polynomials for student age and grade point average, as well as indicator variables for section starting time, student gender, and student nationality. 95 percent confidence intervals based on standard errors clustered at the course level are in parentheses.

**$p < 0.05$.

Heterogeneity by Prior Teaching Experience

A common finding is that the marginal returns to experience diminish over an instructor's career (Papay and Kraft 2015). It may therefore be that inexperienced instructors receive a larger benefit from teaching repetition, relative to more-experienced colleagues. We investigate this possible heterogeneity by stratifying the sample of students based upon whether their instructor is a student (bachelor's, master's, or PhD) or a more senior instructor (postdocs, lecturers, and assistant, associate, and full professors).

Panel A of table 4 shows the effects of teaching repetition in courses taught by students. For these instructors, the effects of repetition on grades, the probability of dropping the course, and study hours appear as before—small and statistically insignificant. Unlike table 3, column 3 suggests economically relevant and statistically significant positive effects of repetition on teaching evaluations for instructors. Students and PhD students receive 16 percent, 24 percent, and 29 percent of a standard deviation higher evaluations in their second, third, and fourth sections, respectively. However, these effects are less precisely estimated than those in table 3, and the F-test for joint significance for all section indicators fails to reject the null hypothesis. We therefore interpret these results as merely suggestive evidence that teaching repetition improves teaching evaluations for instructors who are students.

Panel B of table 4 shows the same estimates for senior instructors. For these instructors, we do not see any evidence that teaching repetition affects student grades,

their probability of dropping out of a course, or their teaching evaluations. There is, however, some indication that being in a senior instructor's second, third, and fourth sections reduces students' study hours. Yet, the point estimates are not statistically significant and we fail to reject the joint significance tests that all section indicators equal zero. Therefore, we are not inclined to see repetition by senior instructors as a relevant factor affecting students' study hours.

Teaching repetition could be particularly valuable for instructors who teach a specific subject for the first time. We test this hypothesis by estimating the main results separately by whether any instructors taught a specific curriculum—as identified by the course code—before. For this specification, we exclude observations from the first year of the dataset for which we do not observe prior teaching experience. Table A.5 in the online appendix shows that these results are qualitatively similar to the heterogeneous results by instructor career experience. There is no evidence of teaching repetition affecting students' grades, dropout probability, or study hours. Although not statistically significant, the point estimates suggest that first-time instructors' teaching evaluations benefit from teaching repetition.[9]

Heterogeneity by Spacing of Repetitions

Are the positive returns to rapid teaching repetition offset by the more general effects of teaching fatigue? In this subsection we investigate whether the spacing of repetitions modulates the effect of teaching repetition. The psychology literature on how practice aids knowledge acquisition in students suggests that the timing and spacing of practice is important (Gerbier and Toppino 2015; Kang 2016). Here, we investigate the possibility that making improvements from repetition requires some short downtime either to reflect on recent experiences or simply to work on implementing pedagogical changes (e.g., reorganizing materials) prior to the next section.

To estimate the effect of repetition spacing, we distinguish whether an instructor had a break—that is, did not teach a section (of the same or different course) immediately before the section under consideration. At this business school, each day consists of five two-hour teaching slots that are separated by thirty minutes to allow instructors and students to change rooms. Instructors who have a break, therefore, have at least two hours and thirty minutes to rest and potentially make changes for their next sections. Empirically, we estimate this effect of section spacing by including interaction terms of a break with second- and third-section indicators (we do not observe a single instance where an instructor had a break before their fourth section).

Table 5 shows the estimates of this fully interacted model. We see no evidence that having a break significantly changes the effects of teaching repetition. None of the eight interaction terms is significant at the 10 percent level. Although these coefficients are less precisely estimated, the direction of the point estimates shows no obvious pattern: Three coefficients suggest having a break increases the benefits of section repetition (e.g., increases grades, lowers dropout rates), and five coefficients suggest the opposite. The F-test for joint significance of all interaction terms does not support the hypothesis that having a break modulates the repetition effect for any of the outcomes we look

---

9. We also estimated the effect of teaching repetition separately for mathematical and nonmathematical courses. In these unreported regressions, we do not see any significant heterogeneity by course type.

Table 5. The Effects of Repetition with and without Break before Section

Dependent Variable	Standard Grade (1)	Dropout (2)	Standard Evaluation (3)	Hours (4)
2nd section	0.000 (−0.052 to 0.053)	0.006 (−0.008 to 0.021)	0.024 (−0.115 to 0.164)	−0.195 (−1.130 to 0.740)
3rd section	0.031 (−0.073 to 0.135)	0.015 (−0.014 to 0.043)	0.017 (−0.250 to 0.284)	−0.576 (−2.244 to 1.092)
4th section	0.046 (−0.100 to 0.191)	0.017 (−0.022 to 0.056)	0.044 (−0.326 to 0.414)	−0.801 (−3.155 to 1.553)
2nd section × break	0.016 (−0.049 to 0.081)	0.011 (−0.009 to 0.031)	−0.076 (−0.226 to 0.074)	0.281 (−0.858 to 1.421)
3rd section × break	0.022 (−0.042 to 0.086)	−0.002 (−0.023 to 0.019)	−0.037 (−0.204 to 0.130)	0.188 (−1.222 to 1.598)
Observations	77,269	83,195	27,144	26,918
R^2	0.569	0.290	0.551	0.412
Section 1 average outcome	.028	.072	−.033	14.415
p-value joint significance of all break interactions	0.7565	.5081	.609	0.8711
p-value joint significance of all section variables + interactions	.6544	.8682	.7510	.8142

Notes: All regressions include instructor-course-parallel-course fixed effects. Additional controls include cubic polynomials for student age and grade point average as well as indicator variables for section starting time, student gender, and student nationality. The reference group is the same as in table 3, that is, students taught in an instructor's first section. 95 percent confidence intervals based on standard errors clustered at the course level are in parentheses.

at. Overall, we interpret these findings as evidence that potential positive effects from repetition are not modulated by short-term fatigue.[10]

5. CONCLUSION

While teaching repetition is pervasive in higher education, we know very little about how it affects teaching effectiveness. Overall, this paper finds evidence that teaching repetition neither hurts nor helps objectively measured teaching effectiveness. Although we find some suggestive evidence that teaching repetition improves teaching evaluations, especially for inexperienced instructors, we can rule out economically meaningful effects of teaching repetition on students' grades, dropout rates, and study hours.

The finding that university instructors' effectiveness is largely unrelated to teaching repetition has a number of implications. First, teaching repetition offers a promising way to reduce overall preparation time that does not harm students. Instructors do not appear to use the first section as a "trial" or "practice run" for later sections, and students in earlier sections are not disadvantaged relative to peers in later sections. A second conclusion is that instructors appear to need significant time to incorporate the lessons from teaching experience.

10. Many other course or instructor characteristics could contribute to heterogeneity in the effects of teaching repetition. For example, there is emerging evidence that students may be particularly critical when evaluating female instructors' teaching performance (Mengel, Sauermann, and Zölitz 2019; Fan et al. 2019), especially in areas involving teaching delivery style and perceived knowledge of the material (Boring 2017). In unreported regressions, we explore whether the effect of teaching repetition on teaching evaluations also differs by instructor gender by estimating models in which we add interaction terms of section order dummies with an instructor gender indicator, but we do not find any statistically meaningful heterogeneity.

Although Williams and Shapiro (2018) find positive effects of teaching repetition on grades, most of their data include classes following a seminar format with a single instructor. The difference in our results suggests that teaching repetition effects may not generalize to the tutorial setting. One reason may be that repetition only improves certain aspects of teaching (e.g., presenting and introducing concepts) but not the skills more applicable to tutorials (e.g., guiding applications). A second reason may be that the effect of tutorial repetition on grades is harder to detect in our setting because all students follow the same lectures, which also affects student grades. This rationale is consistent with our finding that teaching repetition improves evaluations of the tutorial instructor, an outcome that is not directly affected by what happens in lectures. In this light, results from Williams and Shapiro (2018) may better represent teaching repetition effects at small higher education institutions and secondary schools where the single instructor format is more common, whereas our results are more applicable to the lecture-tutorial format that dominates instruction at large higher education institutions.

We may also underestimate the total effect of teaching repetition within our setting if repetition affects all sections of the course similarly, including the first. For example, if faculty prepare more thoroughly for content that they must teach multiple times, then even students in the first sections will benefit from increasing teaching repetition. Here, simply looking at differences across sections within a term will underestimate the effects of teaching repetition on students.

Our estimates begin to reveal how course experience improves teaching productivity in the long run, which is the concern of much of the current research on instructor experience. Given the lack of rapid improvement in teaching from short-term repetition, our results support the idea that teachers need a period of reflection to be able to benefit from their teaching experience. In such a reflection period, teachers can see course evaluations, reflect on experiences, and make substantial changes to the curriculum. To answer how instructor experience translates into better student outcomes, future work should focus on mechanisms that operate on longer time horizons.

REFERENCES

Boksem, Maarten A. S., Theo F. Meijman, and Monicque M. Lorist. 2006. Mental fatigue, motivation and action monitoring. *Biological Psychology* 72(2): 123–132.

Boring, Anne. 2017. Gender biases in student evaluations of teaching. *Journal of Public Economics* 145:27–41.

De Vlieger, Pieter, Brian Jacob, and Kevin Stange. 2018. Measuring instructor effectiveness in higher education. NBER Working Paper No. 22998.

Fan, Y., L. J. Shepherd, E. Slavich, D. Waters, M. Stone, R. Abel, and E. L. Johnston. 2019. Gender and cultural bias in student evaluations: Why representation matters. *PloS One* 14(2): e0209749.

Feld, Jan, Nicolas Salamanca, and Ulf Zölitz. 2020. Are professors worth it? The value-added and costs of tutorial instructors. *Journal of Human Resources* 55(3): 836–863.

Feld, Jan, and Ulf Zölitz. 2017. Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics* 35(2): 387–428.

Gerbier, Emilie, and Thomas C. Toppino. 2015. The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education* 4(3): 49–59.

Harris, Douglas N., and Tim R. Sass. 2011. Teacher training, teacher quality and student achievement. *Journal of Public Economics* 95(7–8): 798–812.

Kaikai, Septimus M., and Regina E. Kaikai. 1990. Positive ways to avoid instructor burnout. Paper presented at the National Conference on Successful College Teaching, Orlando, FL, March.

Kang, Sean H. K. 2016. Spaced repetition promotes efficient and effective learning: Policy Implications for instruction. *Policy Insights from the Behavioral and Brain Sciences* 3(1): 12–19.

Lorist, Monicque M., Maarten A. S. Boksem, and K. Richard Ridderinkhof. 2005. Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research* 24(2): 199–205.

Maslach, Christina, Wilmar B. Schaufeli, and Michael P. Leiter. 2001. Job burnout. *Annual Review of Psychology* 52(1): 397–422.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2019. Gender bias in teaching evaluations. *Journal of the European Economic Association* 17(2): 535–566.

Ost, Ben. 2014. How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics* 6(2): 127–151.

Papay, John P., and Matthew A. Kraft. 2015. Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics* 130:105–119.

Uttl, Bob, Carmela A. White, and Daniela Wong Gonzalez. 2017. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation* 54:22–42.

Williams, Kevin M., and Teny Maghakian Shapiro. 2018. Academic achievement across the day: Evidence from randomized class schedules. *Economics of Education Review* 67:158–170.