International Review of Economics Education 44 (2023) 100271

Contents lists available at ScienceDirect

International Review of Economics Education

journal homepage: www.elsevier.com/locate/iree

Studying like a nerd: Spacing, self-testing, and explanatory questioning in principles of microeconomics

Gabriel X. Martinez [1]

Department of Economics, Ave Maria University, 5050 Ave Maria Blvd., Ave Maria, FL 34142, USA

ARTICLE INFO

JEL classifications:
A2
A22

Keywords:
Science of learning
Spacing
Self-testing
Metacognition
Mastery learning
Academic performance

ABSTRACT

Ineffective study strategies breed overconfidence, bad grades, misplaced blame, and little long-term learning. The psychological learning-science literature suggests that pedagogies based on spacing, self-testing, and explanatory questioning push back against these tendencies and improve learning. This hypothesis was tested by comparing the results in "Chalk-and-Talk" versus "Active Learning" microeconomics courses that used learning science insights. In the latter, the better students were more likely to earn As and middle students' grades averaged about a letter grade higher. Weaker students' failures and Bs were both more frequent. The quality of comments on the readings and lecture completion were highly significant predictors of over-performance for above-median students. Required practice exams and preparation time were significant predictors of over-performance for below-median students.

Although professors want students to finish their courses with a better understanding of the subject, they also want their students to gain skill as learners, growing in responsibility, self-regulation, perseverance, and deferred gratification. Yet many students are ineffective learners (Williamson, 1937, 1939; Tait and Entwistle, 1996; Karpicke et al., 2009; Sauvé et al., 2018). Students struggle with false confidence regarding their knowledge, misplacing blame on the teacher or the test rather than engaging in the self-reflection needed for learning (Chew and Cerbin, 2021).

Learning scientists have made tremendous advances in our understanding of how learning happens in recent decades. There is evidence that, compared to cramming and massing information together, spacing study sessions and materials helps long-term retention of information (Roediger and Pyc, 2012), that repeated self-testing helps solidify knowledge and breeds realistic confidence (Pyc et al., 2014), and that having students explain the material to themselves and to others improves student motivation and facilitates their application of knowledge to new contexts (Dunlosky et al., 2013). However, many economics instructors fail to apply these findings in their courses, continuing to employ a variation of "Chalk-and Talk" pedagogy (see Becker and Watts, 1996, 2001; Watts and Becker, 2008; Boyle and Goffe, 2018; Sheridan and Smith, 2020; Asarta et al., 2021).

This paper describes a systematic pedagogical innovation in line with learning science in five sections of a microeconomics course. It then rigorously compares performance in these Active Learning courses against Chalk-and-Talk courses taught over the same period. The Active Learning courses divided the material into small chunks with strictly enforced deadlines. Students watched short video lectures, used a social annotation tool to elaborate on the reading, checked their learning frequently, and prepared for exams with practice exams. Students explained their knowledge to themselves and others, tested their knowledge, received correction, and developed realistic performance expectations.

E-mail address: gmartinez@avemaria.edu.
[1] https://orcid.org/0000-0003-0707-3520.

https://doi.org/10.1016/j.iree.2023.100271
Received 10 January 2023; Received in revised form 3 August 2023; Accepted 4 August 2023
Available online 6 August 2023

The motivation for the pedagogical innovation is that "the one who does the work does the learning" (Doyle, 2008, 25). Taken as a whole, the system applied the principles of retrieval, spacing, proper foundation, and elaboration explained by Brown et al. (2014). They argue that students learn more when they study frequently rather than cram, when they check their learning rather than assume they know, when they have a good foundation rather than move on unprepared, and when they elaborate on the material rather than merely read it.

The next section summarizes the literature on the active-learning elements used, followed by a detailed description of the courses in practice. The claims are then tested empirically. The result is that the Active Learning system appears to have helped the better students solidify their As, helped middle students do better, raising their median grade by about a letter grade, and caused a bifurcation among weaker students: although their average grade was unchanged, both Bs and failures became more frequent among them. Weak-student bifurcation may be due to the fact that the Active Learning system is designed to push back against procrastination, which may be related to low motivation (cf. Michinov et al., 2011).

1. Related literature on active learning

1.1. Spacing versus cramming

Information retrieval is crucial to long-term memory: reviewing and re-learning previously forgotten material deepens learning (Cepeda et al., 2006). Separating study time into short, frequent sessions – rather than infrequent "cram" sessions – improves sleep patterns, encourages review, and improves memory (Schwartz et al., 2016). To enforce spacing, courses can assign work in small chunks with strict deadlines.

Many papers in economics and other disciplines examine the relation between small-quiz grades and final performance but fail to find a significant relation, possibly because they lack a role for due dates (Galizzi, 2010; Hernández-Julián and Peters, 2012; Maclean and McKeown, 2013; Latif and Miles, 2020). When deadlines are enforced, weekly quizzes are found to improve performance in the final exam relative to a single midterm (Gholami and Moghaddam, 2013). Miller and Schmidt (2021), likewise, find that evenly-spaced deadlines contribute to long-term knowledge retention. Interestingly, Miller et al. (2019) had found that spacing deadlines evenly (as opposed to allowing students to bunch their assignment submissions and turn everything in right before the test) was associated with lower assignment and test scores while Asarta and Schmidt (2013) had found that a self-directed pattern of continuous engagement with the material was significantly and positively related to course performance. The results of these three papers suggest that the underlying determinant of performance may be a student's willingness to adhere to the teacher's system (for example, by accessing materials or completing assignments even if they are not strictly due or even if they can be skipped without further penalty).

1.2. Self-testing versus over-confidence

Students recall significantly more material if repeatedly tested (Roediger and Karpicke, 2006). Self-testing helps retrieval (Roediger and Pyc, 2012; Brown et al., 2014) and builds confidence (Fiorella and Mayer, 2015, 2016). Moreover, practice tests provide the student with an accurate assessment of readiness if they are similar to the graded test in coverage, difficulty and grading standards, and format and length. If they withhold the correct answer after submission, practice tests make students think for themselves. Calimeris and Kosack (2020) find that when the pretest gives the correct answer, there is no significant effect on student learning.

The question of exam overconfidence is well-explored in the literature (for recent examples, see Hossain and Tsigaris, 2015; Yandell and Dirk, 2017; Sawler, 2021). The Dunning-Kruger effect – which suggests that novices tend to overestimate their performance – is the foundation of much of this research (Kruger and Dunning, 1999; Dunning, 2011; Schlösser et al., 2013). Grimes (2002) found that misunderstanding exam coverage contributed to overconfidence in principles of economics. Nowell and Alston (2007) used the midterm to provide midpoint information regarding performance and found that increasing its weight in the grade reduced over-confidence. Henderson et al. (2018) promoted engagement and self-efficacy with a prequiz-learning-postquiz system.

1.3. Explanatory questioning versus passivity

Explanatory questioning (elaborative interrogation and self-explanation) improves memory and comprehension (Roediger and Pyc, 2012; Dunlosky et al., 2013). Elaborative interrogation involves asking whether a statement is true and providing reasons for or against it, perhaps in the context of explaining it to others. Self-explanation involves metacognition: thinking about learning while learning, e.g., planning out learning or monitoring whether it has taken place (Chew and Cerbin, 2021) or rephrasing jargon (Fisher, 1998).

In economics, McGoldrick and Schuhmann (2016) find that grades in a subsequent exam improve if students replace poor grades on a multiple-choice quiz with an essay-based quiz. Sawler (2021) reports that students who discuss the uncertainty of knowledge in economics can predict their grades with increased accuracy. Josephson et al. (2019) found a high correlation between grades in a peer-reviewed essay and final exam scores. Yamarik (2007) asked students to help each other find answers to questions. Emerson et al. (2018) used think-pair-share to bring parity of grades and student satisfaction between large and small courses (Josephson et al., 2019 used it to reduce misunderstanding).

1.4. Proper foundation versus learning gaps

Often "academic ability" is taken to mean the potential level of achievement. The Mastery-Learning approach holds that all students can achieve any desired learning standard, and that "ability" represents the time and effort involved in achieving it. Accordingly, Mastery Learning courses are designed so that assessments are followed by feedback highlighting learning gaps. Teachers use formative assessments to vary teaching speed and methods, including suggesting alternative learning methods or additional practice (Guskey, 2015). By ensuring a proper foundation for subsequent learning, these courses address the belief that mistakes cannot be corrected (Chew and Cerbin, 2021). Millea and Grimes (2002) concluded that disappointed grade expectations influenced student teaching evaluations.

Flipped instruction can aim at mastery learning, with mixed results (Picault, 2019). For example, Balaban et al. (2016), Calimeris and Sauer (2015), Caviglia-Harris (2016), and Wozny et al. (2018) find that it significantly improves learning outcomes, while Olitsky and Cosgrove (2016) find only modest effects and Craft and Linask (2020) find no significant effect.

1.5. Economics teaching and learning science

Economic education research has not studied effective study strategies deeply (Allgood and McGoldrick, 2020, 47). There are exceptions, including Boyle and Goffe (2018), who used a variety of cognitive strategies to improve their principles of macroeconomics course and found an effect size of 0.8 standard deviations and Cosgrove and Olitsky (2020), who built upon learning-science principles and found 10–12 % point gains in grades.

2. Applying active learning in a microeconomics course

An active-learning philosophy shaped five 2020–2021 sections of an introductory microeconomics course. Students learned through videos with embedded quizzes, readings that required elaboration, and self-checks, all prepared by the same instructor. They received feedback and re-read or re-watched to find the correct answers and could not move on to the next item without passing the previous item. The system was designed to work as a coherent whole. In the words of one of the students,

> The design of the course was very smart. You watched lectures, commented on reading assignments, and then took small quizzes that can be taken as many times as needed. This allows you to answer test-like questions to get an idea of what it was like, but it wasn't too stressful. This gave you an idea of what the questions would be like and as much practice as you like.

Active Learning courses subdivided twelve chapters into 2 or 3 sections each. Table 1 displays how the work was spread out over a sample week.[2]

Fig. 1 illustrates the course structure, with a few items broken down for illustration. The course comprised three units of four chapters subdivided into sections. Each section contained a ten-minute video lecture, a five-page reading assignment, and a ten-question formative assessment. These could be reviewed and repeated without limit. After two chapters, students encountered two practice quizzes and a graded quiz. Each unit closed with a four-chapter test. The following sections explain each item in more detail and show the connection of each assignment with the principles of spacing, self-testing, explanatory questioning, and proper foundation.[3] Assignment instructions roughly follow the "task, purpose, criteria" outline of transparent teaching (Winkelmes et al., 2016) described in Berrett (2015): instructions emphasize the importance of each assignment for acquisition of knowledge and skills and the attainment of course learning goals; they describe what it is to be done and how; and they show what excellence looks like (often with annotated examples) and provide standards for self-evaluation.

2.1. Video lectures

The syllabus instructed students to watch the videos (and the Canvas progression enforced this instruction), taking notes and paying attention. Panopto video lectures preceded reading to contribute to proper foundation. The deadlines (and a 4-percent late penalty) encouraged spacing. In-video quizzes provided self-testing every few minutes, aiming at ensuring attention and basic understanding (the average grade was 98 %). Progress in the video requires completion of the in-video quizzes (placed at frequent intervals through the videos). Embedded quizzes also "chunked" the material to reduce the load on working memory (Goffe, 2021). [4] In the words of one of the students,

[2] The organization of the assignments was the same across sections (in the sense that the order of the material and enforced order of assignments is identical). For most students, the time available for completion was very similar (Summer courses take half the time, but Summer students take half as many courses at one time). In the regression, inclusion of a "term" indicator variable did not change the results, and the term dummy was not significant. Both pedagogies (Active Learning and a more standard pedagogy) were used in both Summer and Fall.

[3] Additionally, four assignments asked for a news-article analysis to "promote reasoning and application in novel, but still similar, contexts" (Chew and Cerbin, 2021, 12).

[4] https://www.panopto.com/. Some students fast-forwarded to the quizzes and skipped the lectures. The variable Avg Video Completion below captures the information value of this behavior.

Table 1
Sample weekly course schedule.

	Monday	Tuesday	Wednesday	Thursday	Friday
Summer	Lecture 8.1	Lecture 8.2	Lecture 8.3	Lecture 8.4	First Practice Test 2
(7.5 weeks)	Reading 8.1	Reading 8.2	Reading 8.3	Reading 8.4	2nd Practice Test 2
	Check 8.1	Check 8.2	Check 8.3	Check 8.4	Test 2
Fall	Lecture 8.1	Check 8.1	Lecture 8.2	Check 8.2	
(15 weeks)	Reading 8.1		Reading 8.2		

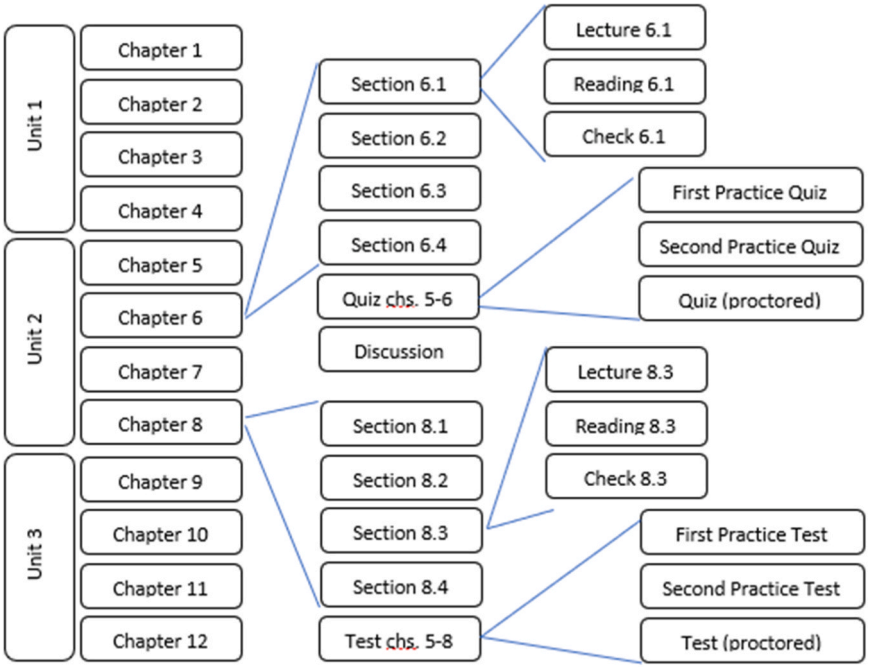

Fig. 1. Course structure. Access to each item depended on the completion of previous items.

> I liked [the class] a lot. There were video lectures to watch in advance of class. I liked this method because I could stop and go through the lecture or go back.

The summer online course featured no other regular instruction (the instructor was continuously available). Fall on-campus students were able to access the teacher during class time; class time was devoted to answering questions.

2.2. Collaborative readings

The course textbook consisted of detailed lecture notes written by the instructor and available on the Perusall social annotation platform.[5] Fig. 2 shows how students commented on the text or responded to others' comments. Students are told that they would earn points on account of the quantity and quality of their comments (high-quality, insightful summaries; interesting, thought-provoking questions) and by collaborating with others. Comments are graded by a proprietary AI tool, so that grades aren't given for accuracy but for thoughtfulness. The instructor has to check that the students are not gaming the system, and there are multiple ways to check.

The requirement to make substantial, detailed comments to encourage explanatory questioning. Comments that asked for (or offered) clarification or correction served as a self-test. Spacing was encouraged by strict late penalties, while the video-then-reading-then-check sequence built a proper foundation. Grades varied widely around a mean of 80 %: the standard deviation was 22 %.

[5] Based loosely on a well-known textbook used in C&T courses at the university. Perusall is free of charge https://perusall.com/. Students access the reading, highlight part of it, and make comments on the side. They can respond, tag, and upvote each other. For auto-grading details, see https://perusall.com/downloads/scoring-details.pdf. For examples, see https://perusall.com/downloads/scoring-examples.pdf

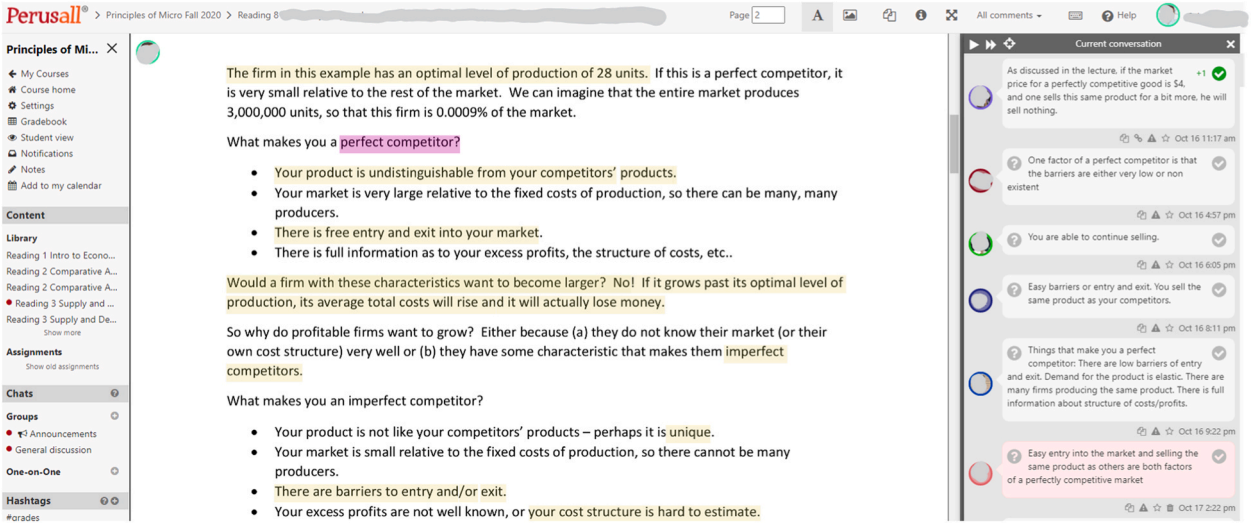

Fig. 2. Sample Perusall assignment.

2.3. Learning checks

Each chapter section ended with a knowledge check: ten auto-graded questions in Canvas New Quizzes, based on the lecture and reading.[6] Students were given these instructions at the start of every Check:

> Watch the video and do the reading. Then answer these questions. Being a take-home check, it is open-book, open-notes; because it covers the same topics as the Quiz and Test, it is a good preparation for them. You want to take it as seriously as you would a real exam.
>
> In order to pass this quiz, you must score at least 80 %. You will be allowed unlimited attempts. After each attempt, you will be able to see your score and whether your answers were correct. You will be asked to wait 10 min before a new attempt – use this time to review what you got wrong, re-watch the lecture, and re-read the reading.
>
> A penalty will apply if you do it late – do it early in the day! Get ahead! Give yourself plenty of time to get it right.
>
> These policies, I have to admit, are a little annoying. Why check your knowledge? Why a minimum grade? Why the 10-minute waiting period? Why the late penalty? The short answer is: I want you to learn. It's easy to glide through a class. It's tempting to coast and cram. But if you do, you don't learn anything. These policies discourage the coast-and-cram strategy and make sure that you actually learn something.

Self-tests were due at the end of each chapter section. Students could ascertain the correct answers by re-accessing the adjacent reading and video. They often re-took the check to earn a perfect score: the average grade was 92 %, with a standard deviation of 4.5 % points.

Learning Checks were the main self-testing component. Enforcing a minimum standard (a minimum grade of 80 %) built a proper foundation and confidence in future success. The late penalty (4 % per day, up to 20 %) helped space knowledge.

2.4. Quizzes and tests, graded and practice

Quizzes (tests) covered two (four) chapters. Exams were not explicitly cumulative. Two practice exams (one required, one optional) preceded each graded exam. Questions were drawn from the same test bank and had the same time limits and length. While graded exams had fixed questions, practice exams randomly selected questions for each submission.

Practice exams were open-book, allowed unlimited attempts, kept the highest grade, and revealed whether a question had been answered correctly upon completion. Graded exams were closed-book, could only be taken once (on a locked browser), revealed only the overall score, and were proctored in person or with a webcam recording service.[7] As self-tests, practice exam could be repeated (and used to identify lacunae in understanding) up to the desired level of performance, re-accessing adjacent readings and videos.[8] The gate-keeping and late penalty played a role in spacing. Students expressed appreciation for being able to build a proper foundation and confidence (as well as occasional frustration). In the words of one of the students,

> The best thing about this course was the practice exams before the exam. These are very helpful in studying for the exam. It covers all the material that will be on the exam in detail.

Students needed a minimum score of 71 % in the required practice exam to take the graded exam. Fig. 3 shows that many students stayed close to this minimum requirement. The vast majority of students did not improve their scores with further attempts once they had achieved a "passing" score, even though practice scores predicted test scores rather well.

Suppose students had taken full advantage of the opportunity to practice. Practice scores would not be able to predict graded exam scores in a regression, as perfect practice scores would be the norm. The number of practice attempts, likewise, would be uncorrelated with graded exam scores: students would achieve complete learning by their last attempt, whether the first or the seventh.

There is evidence that students did not take full advantage of the formative opportunity. Table 2 reports that the relation between practice scores and graded-exam scores was positive: controlling for GPA, a student who scored a standard deviation (ten percentage points) better on the practice exams scored three points better in the graded quizzes and eight points better in the graded tests. But the relation between practice attempts and graded exam scores was negative: students with similar GPAs who took two extra attempts[9] to achieve the same practice-exam score scored worse in the quizzes and tests. A reasonable interpretation is that weaker students took the practice exam repeatedly until they achieved the minimum score required to take the graded exam, and reverted to mean after a

[6] For information on this quizzing format, see https://community.canvaslms.com/t5/Instructor-Guide/How-do-I-create-an-assessment-using-New-Quizzes/ta-p/1173

[7] LockDown Browser https://web.respondus.com/he/lockdownbrowser/. Webcam monitoring https://web.respondus.com/he/monitor/

[8] Students took ninety (eighty) percent of the optional practice quizzes (tests). The mean number of second practice quiz (test) attempts was 1.5 (0.8). Scores and the number of attempts in optional practice exams are not significantly related to scores in for-a-grade exams.

[9] The standard deviation of the number of attempts for practice quizzes was 2.21 (for practice tests, it was 1.85). This makes sense, as quizzes were half as long as tests.

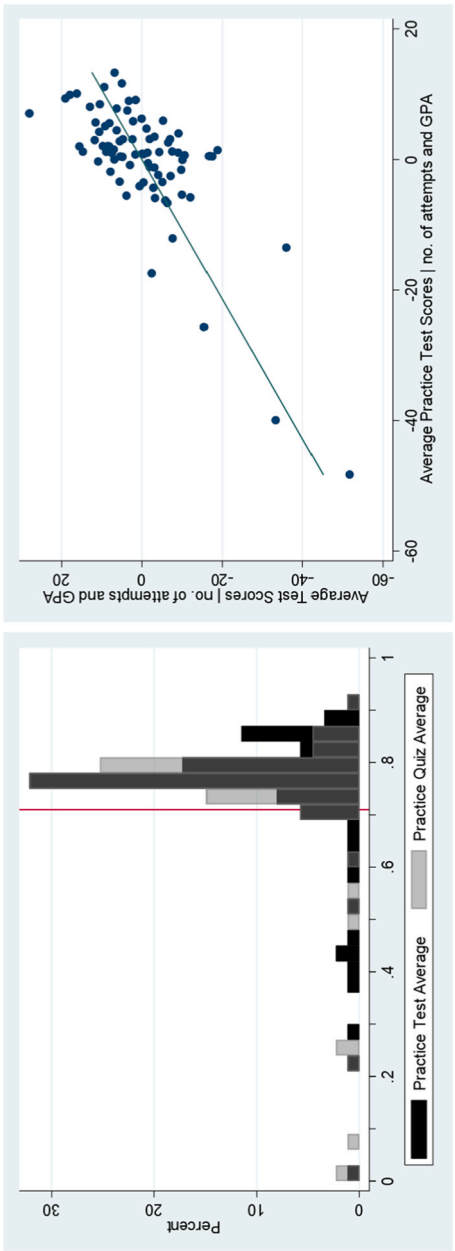

Fig. 3. Distribution of practice scores and partial correlation plot of average test scores against average practice test scores, controlling for number of attempts and GPA.

Table 2
Predicting quiz and test scores with practice exam scores and number of attempts, controlling for GPA. Standard errors in parentheses, * $p < 0.05$, * * $p < 0.01$, * ** $p < 0.001$.

	Quiz Average Scores			Test Average Scores		
Practice Avg	0.408**	0.485***	0.320*	0.857***	0.935***	0.807***
	(0.130)	(0.131)	(0.124)	(0.114)	(0.104)	(0.0964)
Attempts		-1.490*	-1.230*		-2.550***	-2.269***
		(0.629)	(0.572)		(0.583)	(0.521)
GPA			9.027***			7.602***
			(2.116)			(1.624)
Constant	42.30***	42.30***	25.77**	7.799	10.22	-4.734
	(10.04)	(9.758)	(9.644)	(8.672)	(7.840)	(7.661)
Observations	81	81	81	81	81	81
Adjusted R^2	0.099	0.149	0.303	0.411	0.521	0.622

fortunate draw. Moreover, because these effects are fairly robust to controlling for GPA, practice exams appear to help low-GPA, hard-working students narrow the gap relative to high-GPA students.

2.5. Course progression

By using Canvas module prerequisites and enforcing sequential order of requirements, the course required students to submit every assignment in order (see Fig. 4). A few students fell behind and attempted many assignments immediately before the test. The strong alignment between learning materials, practice exams, and graded exams forestalled blaming the teacher or the tests. [10]

3. Empirical analysis

3.1. Hypothesis

Do students with comparable aptitude, motivation, and preparation perform differently in Active Learning (AL) courses versus Chalk-and-Talk (C&T) courses? To answer this question, an active learning strategy was implemented in some sections of a micro-economics course while a Chalk-and-Talk methodology was used in other sections. Fig. 5 indicates that raw performance in AL courses appears better than in C&T courses: more frequent As and Bs and fewer Cs and Ds, although the proportion of failures and withdrawals is substantially higher.

The AL strategy, described above, should be unnecessary for students with good study habits, beneficial for those with shortcomings who are willing to work, and the cause of withdrawal for those who want to coast and cram. This section tests this hypothesis using Canvas learning management system (LMS) data on course performance and academic records from the student information system, comparing performance in five AL sections against performance in eight C&T sections. C&T sections are taught following a standard methodology. They.

- meet in a classroom for 200 min a week;
- deliver information through in-class lectures, with some question-and-answers opportunity;
- schedule textbook reading;
- require (weekly or biweekly) homework through an online portal provided by the publisher;
- have flexible deadlines;
- require quizzes at the end of each chapter through the textbook portal;
- include high-stakes tests drawn from a publisher-provided test bank; and
- cover the same twelve chapters of a popular textbook.

As is common practice, C&T sections did not enforce a specific course progression and set no minimum-performance requirements for advancement to the next portion of the course. Although there was some variation across instructors and semesters, in general C&T sections had no mechanism for enforcing appropriate spacing of work, provided minimal self-testing in the form of study guides or optional practice exams, and did not require any kind of reflection upon the reading.

The book used in all C&T sections was the basis for the instructor-written material used in the AL courses. AL sections based their exams on the C&T test bank (tests and quizzes remained constant, but plagiarism-protected, in Instructor 3 courses). Sections met at similar times (Tuesday-Thursday late morning, Monday-Wednesday-Friday early afternoon, or online in the summer).

Instructors 2 and 3 are full-time economics faculty with Ph.D.s; Instructor 1 is an adjunct professor with a Master's degree in economics. Instructor 2 is a trained microeconomist who taught multiple sections of microeconomics over 2004–2006 and who started

[10] The Dean was made aware of the strategy at an early stage. Hard data about student under-performance was helpful to Student Support and the Athletic Department.

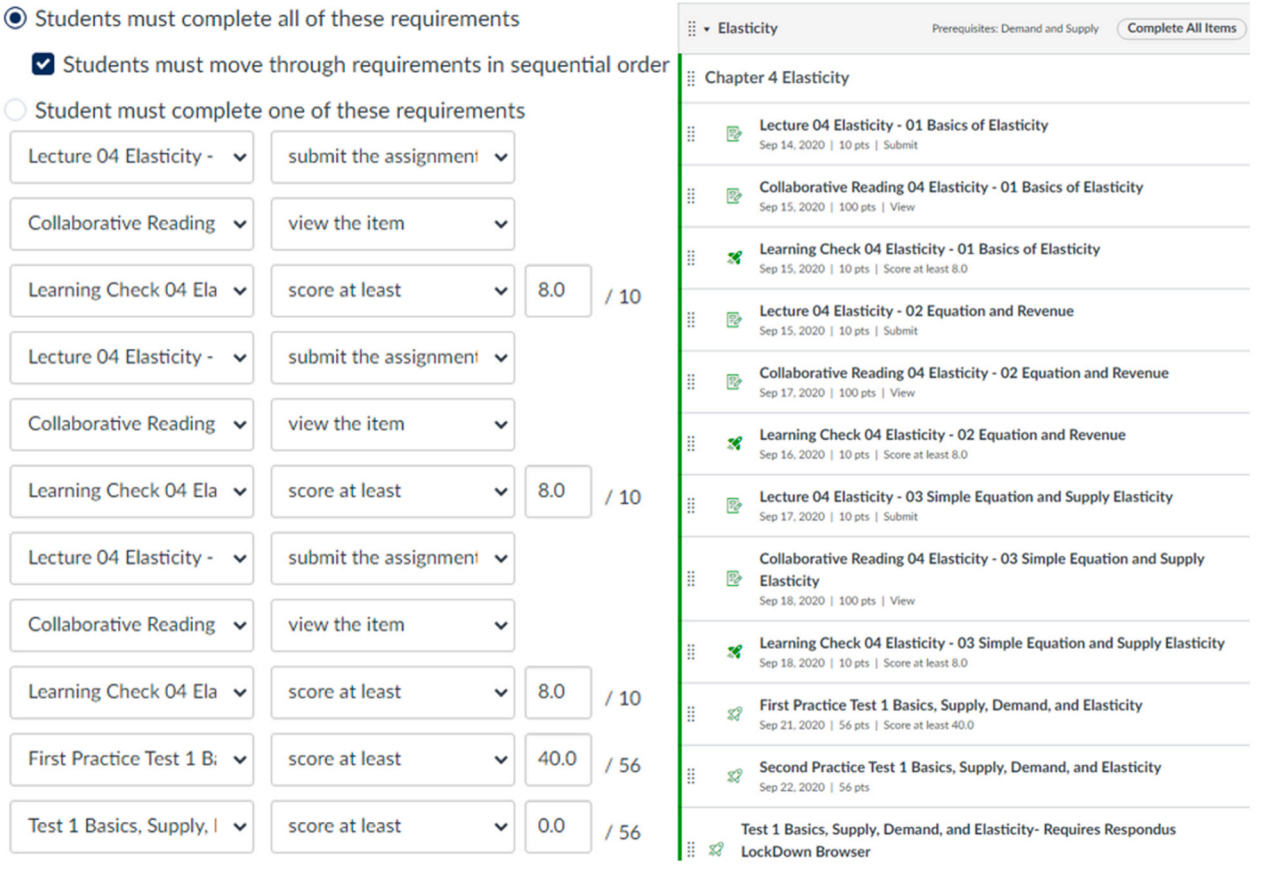

Fig. 4. Sample course progression.

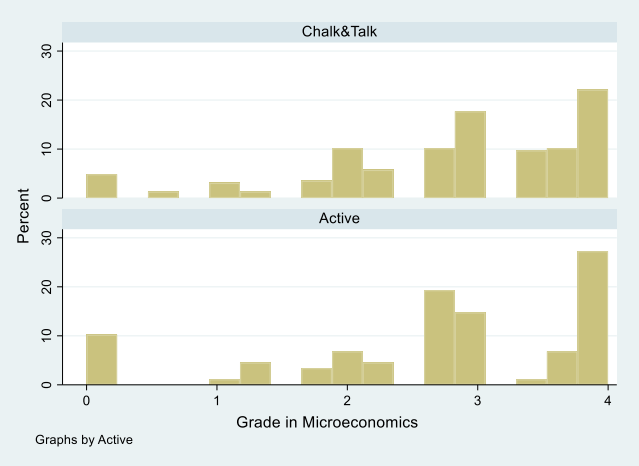

Fig. 5. Raw comparison of grades between C&T and AL courses.

teaching at this institution in 2018; Instructor 3 is a macroeconomist with eighteen years of teaching experience at this institution who had last taught microeconomics in 2003.

Table 3 reports the distribution of students across instructors and course styles, their Microeconomics grades, and the GPAs they had earned up to the semester before taking the course. Table 4 describes the grading schemes of C&T and AL courses.

Five pre-pandemic C&T sections are included in the sample. The Spring 2020 C&T section was interrupted by COVID-19. Fall 2020 performance in both Active Learning sections, as well as the Spring 2021 C&T section, may have been affected by (minor) social-distancing measures in place. Fall 2021 AL sections were taught without COVID-19 interruptions as, by Summer of 2021, sections were back to a pre-pandemic mode. In regressions, indicator variables for those terms (or for those years, or for the relevant instructors) are not significant and their inclusion does not change the result.

3.2. Characteristics of the sample

This is an 1100-student, primarily undergraduate, private, faith-based institution in the southeastern United States. About 60 % of microeconomics students pursue a major in fields related to business or economics, and for them the course is required. 11 % are economics majors. Nine percent of students are freshmen, 46 % are sophomores, 30 % are juniors, and 15 % are seniors.

Table 5 shows that students in AL courses and in C&T courses were not significantly different along most relevant covariates (although AL students tended to be slightly less prepared and display less academic aptitude than C&T students). Although there was no attempt to set up a rigorous randomized controlled trial, students' choice of section was not likely to be determined by the differences in pedagogy.

- Students signed up for Fall 2020 courses in March; the change in course style was implemented in July and no special announcement was made. The same two teachers taught both sections, at the same times and days as in Fall 2019. Thus students likely made their selection of section with the same criteria they had used a year before (e.g., instructor preference, schedule, etc.).
- Fall 2021 sections were originally scheduled to be taught by Instructor 2 when students registered for the course in March. A personal emergency led to a change of instructor during the summer, and he change in course style was not announced. There is a hearsay report of a student dropping the Fall 2021 AL section because of the frequency of assignments, but because the change was made after students' schedules were set, this factor is likely to play a minor role.

Insofar as it is expected that the system will have different impact on different students, it is useful to divide the "control" group of 226 C&T students into quintiles according to their course preparation. Dividing into quintiles strikes a balance between fine-grained detail and a sufficient sample in each sub-group, although similar results are observable with dividing into more or fewer groups. GPA cut-offs for each group are 2.64, 3.42, 3.6, and 3.85. The same cut-offs were then applied to the "treatment" group, the Active Learning courses. Because AL students were slightly less prepared, these sections had proportionately more students in the bottom two quintiles (25 % and 22.73 %, respectively) and fewer high-preparation students. However, as Table 5 shows, failure rates were not significantly different across course styles.

3.2.1. Variable specification

Anderson, Benjamin, and Fuss (1994) find that GPA and prior economics grades (as well as grades in mathematics and English, years at the university, and age and gender) predict success in introductory economics. This paper defines "success in microeconomics" in two different ways.

Table 3
Distribution of sections by instructor, term, and course style.

	Instructor 1			Instructor 2			Instructor 3						Overall		
	Chalk and Talk									Active					
	Obs.	Grade	GPA	Obs.	Grade	GPA	Obs.	Grade	GPA	Obs.	Grade	GPA	Obs.	Grade	GPA
2018 Fall	28	3.21	3.34										28	3.21	3.34
2019 Spring	48	2.84	3.18										48	2.84	3.18
Summer							7	2.67	3.05				7	2.67	3.05
Fall				26	2.75	3.12	20	2.43	2.91				46	2.61	3.03
2020 Spring				35	2.97	3.29							35	2.97	3.29
Summer										18	2.06	2.73	18	2.06	2.73
Fall				30	3.03	3.08				27	2.82	3.06	57	2.93	3.07
2021 Spring				32	2.58	2.94							32	2.58	2.94
Summer										4	3.18	3.59	4	3.18	3.59
Fall										39	2.89	3.08	39	2.89	3.08
	76	2.98	3.24	123	2.84	3.11	27	2.49	2.95	88	2.71	3.03	314	2.81	3.11

Table 4
Alternative Grading Schemes.

	Chalk and Talk		Active Learning
Three Tests	25 %	Three Tests	25 %
Three Quizzes	20 %	Three Quizzes	15 %
Four applications of microeconomics to a news item	15 %	Four applications of microeconomics to a news item	15 %
Optional Practice Exams	4 %	Practice Exams	4 %
		Video Lectures	15 %
Adaptive Reading without elaboration	20 %	Collaborative Reading with elaboration	15 %
Homework without strictly enforced deadlines	20 %	Learning Checks with strictly enforced deadlines	15 %

Table 5
Summary statistics, by course style. Standard deviations in brackets, standard errors in parentheses, + p < 0.1, * p < 0.05.

	All students				Only students who passed			
	Active	C&T	Difference		Active	C&T	Difference	
Uniform Course Grade	76.76	74.5	2.27		81.72	77.02	4.7	*
	[18.44]	[16.79]	(2.18)		[8.25]	[12.44]	(1.51)	
Test Average	68.68	66.83	1.85		74.21	68.94	5.27	*
	[21.2]	[21.39]	(2.69)		[11.34]	[19.13]	(2.29)	
GPA	3.03	3.14	-0.11		3.12	3.18	-0.06	
	[0.66]	[0.61]	(0.08)		[0.61]	[0.57]	(0.08)	
Macro grade	2.74	2.86	-0.12		2.95	2.92	0.02	
	[1.15]	[1.07]	(0.14)		[0.99]	[1.03]	(0.13)	
Math & Stat avg grade	2.97	3.18	-0.22	*	3.08	3.22	-0.14	
	[0.92]	[0.82]	(0.11)		[0.87]	[0.8]	(0.11)	
Dummies								
Previous Macro taken	0.45	0.6	-0.14	*	0.47	0.62	-0.15	*
	[0.5]	[0.49]	(0.06)		[0.5]	[0.49]	(0.06)	
Freshman	0.08	0.1	-0.02		0.08	0.1	-0.02	
	[0.27]	[0.3]	(0.04)		[0.27]	[0.3]	(0.04)	
Female	0.33	0.36	-0.03		0.33	0.36	-0.03	
	[0.47]	[0.48]	(0.06)		[0.47]	[0.48]	(0.06)	
Dummies for Majors								
Economics major	0.09	0.12	-0.02		0.1	0.12	-0.02	
	[0.29]	[0.32]	(0.04)		[0.3]	[0.32]	(0.04)	
Econ-Glb Aff-Mgrl Ec major	0.22	0.19	0.02		0.23	0.2	0.03	
	[0.41]	[0.4]	(0.05)		[0.42]	[0.4]	(0.05)	
Mathematics major	0.01	0.03	-0.02		0.01	0.03	-0.02	
	[0.11]	[0.16]	(0.02)		[0.11]	[0.17]	(0.02)	
Busn-Finc-Acct-Mktg major	0.64	0.58	0.06		0.66	0.59	0.07	
	[0.48]	[0.49]	(0.06)		[0.48]	[0.49]	(0.06)	
Percentage of failures by Quintile								
Failures in Quintile 1	34.8	15.2	19.57	+				
	n = 23	n = 46	(11.3)					
Failures in Quintile 2	5.3	8.9	-3.63					
	n = 19	n = 45	(6.7)					
Overall	10.2	4.9	5.4	+				
	n = 88	n = 227			n = 79	n = 215		

- **Average test scores** (midterms and final exams). Summative tests were the most comparable element across sections: similar in style, coverage, source, number, and overall weight, and administered using strict anti-plagiarism protocols (a lockdown browser plus webcam monitoring or proctoring).
- The overall grade posted on the student information system is not necessarily a consistent measure of student performance: the weight of homework, quizzes, and tests (and other components, such as attendance or videos) varied over courses. Moreover, "curving" at the end of the semester may have taken place. A **uniform overall grade** is calculated as the average of Tests, Quizzes, Homework, and Reading+Video assignments, with the same weights across courses. A scaling factor adjusts for the difference between the maximum possible percentage and the sum of the weights of assignments excluded in the calculation (e.g., attendance in some sections, discussion boards or extra credit in others). Variations of the formula yielded similar results.

$$uniform \quad grade = (0.2 \times QUIZ + 0.2 \times HW + 0.25 \times TEST + 0.2 \times READVIDEO) \times (\max \quad points \quad possible - points \quad excluded) \times 100$$

Regarding covariates (whose summary statistics are in Table 5),

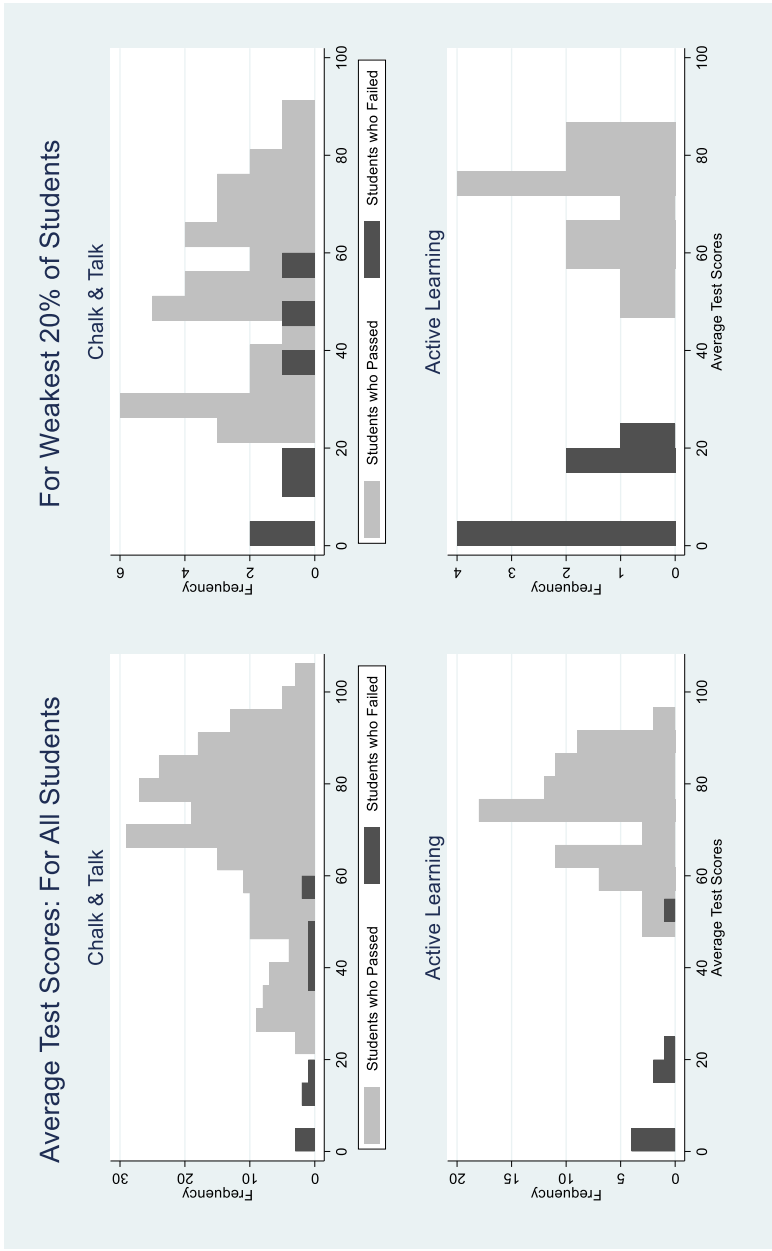

Fig. 6. Distribution of test averages over course styles.

Table 6
Predicting Test Scores in Chalk-and-Talk versus Active Courses. Robust clustered standard errors in parentheses, $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Dependent Variable	Test Score Average							Uniform Course Grade						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
GPA	12.09***	6.408+	12.09***	11.55***	11.51***	11.51***	11.47***	10.41***	8.295***	10.41***	10.19***	10.16***	10.15***	10.12***
	(1.878)	(3.031)	(1.840)	(2.049)	(2.062)	(2.059)	(2.072)	(1.119)	(0.870)	(1.096)	(1.530)	(1.546)	(1.546)	(1.561)
Active L x GPA			-5.686	-3.667	-3.746	-3.967	-4.023			-2.116	-1.204	-1.234	-1.238	-1.255
			(3.394)	(4.997)	(4.943)	(3.756)	(3.784)			(1.368)	(2.344)	(2.284)	(2.244)	(2.202)
Macro grade	5.032*	2.854	5.032*	4.623*	4.661*	4.456*	4.496*	2.415	0.288	2.415+	2.287+	2.334+	2.195+	2.243+
	(2.086)	(2.248)	(2.044)	(1.754)	(1.780)	(1.626)	(1.655)	(1.365)	(0.638)	(1.338)	(1.200)	(1.203)	(1.163)	(1.166)
Active L x Macro grade			-2.178	-2.976	-3.243	-3.332	-3.563			-2.127	-2.429	-2.611	-2.434	-2.599+
			(2.942)	(3.212)	(3.164)	(2.529)	(2.510)			(1.466)	(1.522)	(1.499)	(1.437)	(1.435)
Constant	15.72	45.79***	15.72+	13.59	13.86	13.43	13.73	36.81***	54.97***	36.81***	42.84***	43.15***	42.84***	43.16***
	(8.529)	(4.272)	(8.356)	(11.03)	(11.21)	(11.17)	(11.35)	(4.819)	(2.888)	(4.721)	(7.409)	(7.531)	(7.485)	(7.608)
Active Learning			30.07**	38.47***	38.68***	40.15***	40.28***			18.16**	17.98**	18.02**	18.44**	18.43**
			(9.272)	(5.972)	(5.636)	(6.589)	(6.245)			(5.447)	(5.731)	(5.510)	(5.937)	(5.653)
Instructor, Year, and Term dummies				Yes	Yes	Yes	Yes				Yes	Yes	Yes	Yes
Freshman dummy					Yes		Yes					Yes		Yes
Dummy for Econ-related major						Yes	Yes						Yes	Yes
N	215	79	294	294	294	294	294	215	79	294	294	294	294	294
Adjusted R^2	0.338	0.294	0.341	0.533	0.534	0.537	0.538	0.401	0.397	0.417	0.481	0.482	0.479	0.480
Bayesian Info Criterion	1803.6	590.4	2423.5	2343.8	2351.9	2350.3	2352.9	1596.9	527.7	2147.0	2134.1	2142.8	2144.0	2147.0
RMSE	15.57	9.530	14.22	11.98	11.96	11.92	11.91	9.624	6.406	8.889	8.388	8.380	8.397	8.389

- **Academic ability** is proxied by the student's GPA in the semester prior to taking the course. Also considered: the number of courses the student has repeated, whether the student had progressed beyond algebra or taken accounting, and grades in Math or English. High school GPA and standardized scores were also considered but data availability is more limited because not all students report these (e.g., transfer students, test-optional students).
- **Preparation** is proxied by the grade in Macroeconomics. Macro grades are either the earned grade or an imputed grade on the basis of students' GPA, class level, and grades in math courses, using Stata®'s multiple imputation routine (if the student had not taken the course). Also considered: having taken a macroeconomics course (59.5 % had taken it), whether the student had taken the required math class (97.7 % had), the number of math courses taken, and the number of transferred credits.
- **Maturity** is measured by whether the student is a freshman. Also considered: class level and the number of credits earned at the time of taking the course.
- **Motivation** is proxied by the student's choice of major (economics, economics/global affairs/managerial econ; business administration, business/finance/accounting/marketing; mathematics) and whether they had a second major (7.3 % of students). About 58 % of students were majoring in the business-related programs; 19 % in economics-related programs, 3 % in mathematics, and the rest in one of 19 other programs.
- Dummies for instructor, year, and term were included.

3.3. Empirical model

The estimated model is

$$\begin{aligned} Micro \quad performance_i = {} & \beta_1 \times ability_i + \beta_2(ability_i) \times AL_i + \gamma_1 \times preparation_i + \gamma_2(preparation_i) \\ & \times AL_i + \alpha + AL_i + \delta_1(instructor_i) + \delta_2(year_i) + \delta_1(term_i) + \psi_1(maturity_i) + \psi_2(motivation_i) + \epsilon_i \end{aligned} \tag{1}$$

The uniform overall grade outcome variable and the average test scores outcome variable are predicted using OLS with robust standard errors, clustered over thirteen course sections to account for section heterogeneity.

Asarta and Schmidt (2013) provide evidence for the importance of continued engagement in a course. As Michinov et al. (2011) substantially document, students who procrastinate often lack motivation. The AL system required would-be procrastinators to develop the motivation to meet deadlines. Some students did not follow the system – those students failed and therefore the regressions exclude them. Therefore the correct population for the study is the students who adhered to the protocol: those who persisted through the end of the class.

Put differently, one of the expected consequences of the Active Learning system is to cause a bifurcation of grades for the weaker students. Some students, even very weak students, will benefit from the AL system and obtain a higher grade than they would have otherwise; others will be unable or unwilling to space, self-test, and reflect as required and will ultimately fail.

The bifurcation of grades in AL courses can be clearly seen in Fig. 6.

This bifurcation means that if the regressions included students who ultimately failed the course, the intercept would be biased downward as it would include students who did not, in fact, receive the full treatment. Indeed, in the full dataset, the AL coefficient is smaller and statistically insignificant, with the interpretation that (because the AL coefficient is the intercept of the regression) the very worst students were predicted to score, on *average*, equally badly on Tests on both AL and C&T courses. What Fig. 6 suggests is that the performance of the AL system should be judged on the performance of the students who stayed in the AL system through the end of the semester. For this reason, the regressions below exclude students who ultimately failed the class.

4. Results

Table 6 below reports the results of the regressions. Columns 1–7 have Test Score Average as the dependent variable; the Uniform Course Grade is the dependent variable in columns 8–14. Columns 1 and 8 only include C&T students; columns 2 and 9 include only AL students. A very large difference in the intercept is noticeable, indicating that the least-prepared students (who passed) were expected to do substantially better in AL courses than in C&T courses. This is tested formally in columns 3 and 10 by the coefficient on the Active Learning variable. This coefficient is large, significant, and robust across specifications (columns 4–7, 11–14).

The smaller (and less significant) GPA and Macro grade coefficients in columns 2 and 9, relative to columns 1 and 8, suggest that prior preparation may have been less significant of a determinant of grades in Active Learning courses. Weak students who persisted in the system seemed to have narrowed their gap relative to better-prepared students. The interaction terms in columns 3 and 10, however, are not statistically significant.[11]

The joint hypothesis that the coefficients on the dummy variable and the interaction term are jointly zero can be rejected for both Test averages ($F(2, 277) = 21.23$, $P = 0.0001$) and course grades ($F(2, 277) = 7.44$, $P = 0.0007$). The principal result is that prior performance (as indicated by GPAs and grades in previous courses) seems much less important for course performance in the Active Learning system than in Chalk-and-Talk courses. The difference in test averages between the weakest students and the strongest students is nearly 25 points in Chalk-and-Talk courses and just over 12 points in Active Learning courses.

[11] Following a referee's suggestion, a Class Size variable was added to the regressions with and without dummy variables, as it might have an effect on performance. The slope coefficient on this variable turned out to be insignificant; none of the relevant coefficients changed.

4.1. Implications of the results

The philosophy of Mastery Learning suggests that, as long as students are given feedback and the opportunity to improve, their prior preparation should matter relatively little, and that differences in "ability" should not be interpreted as different caps on performance, but as differences the amount of time (or tries) it will take for two students to perform at the same level.

Fig. 7, produced with Stata®'s *marginsplot* command, summarizes the findings. The blue solid line represents students' performance in C&T sections and compares performance of each quintile against the "middle" quintile. Weaker students (in the bottom and second quintiles) perform significantly less well than middle students (16 and 6 points worse, respectively; the P-values of the corresponding F statistic of the Test Average contrasts are 0.0012 and 0.0415). Students in the fourth quartile do not perform significantly better than the middle-quintile Test average (5.5 points better, but P = 0.2838). Top-quintile students perform better than middle students (17.4 points better, P = 0.0002).

Performance in AL courses is represented by the red dashed line. Bottom-quintile AL students performed just as well as middle students (2.15 points better, P = 0.4925) and significantly better than bottom-quintile C&T students (17.76 points, P = 0.028). Second quintile students performed a little worse than the middle, but not significantly so (−1.1 points, P = 0.8216). The dispersion of performance narrows significantly for fourth quartile students who, unlike in the C&T sections, perform significantly better than middle students in AL sections (11.95 points better, P = 0.0001). Top students also perform very well in AL sections (18.6 points, P = 0.0001). A natural interpretation is that the AL system encourages students to study like the best students.

The right hand side panel of Fig. 7, which shows the differences in performance in overall course grades, does not show an improvement for each of the quintiles. However, it does show a significant improvement in overall grades for the bottom quintile (differences for other quintiles are not statistically significant). It also suggests that the improvement in performance is not likely attributable to across-the-board grade inflation. A plausible interpretation is that the AL system is most helpful as a test-preparation system.

4.2. Active learning and the pandemic

The Active Learning strategy was implemented a few months after the start of the COVID-19 Pandemic. How do these two events interact? Columns 3 and 10 do not include dummies for Year (or other dummies) while Columns 7 and 14 do. Most coefficients are not much changed by the inclusion of Year terms and none of the dummies are significant if the Course Grade is the dependent variable.

On the other hand, if the Test Score Average is the dependent variable, the Active Learning coefficient is larger (by about a letter grade) when the Year dummy is included. With 2019 as the base year, the coefficient of Year 2020 is − 13 (standard error = 0.63) and the coefficient of Year 2021 is − 27 (standard error = 9.26). That is, adding the Year dummies to the Test Score Averages regressions gives evidence of the negative impact of the Pandemic while at the same time increasing the size of the Active Learning coefficient. A tentative interpretation is that the negative impact of the Pandemic on test scores was more than offset by the Active Learning strategy.

4.3. Explaining performance in active learning courses

Which facet of AL courses was most helpful? Is the relative success of the AL strategy merely due to a greater investment of time – to "busywork"?

4.3.1. Empirical strategy

To answer this question, the following OLS regression model is estimated:

$$AL \quad Performance = \beta_0 + \beta_1 X + \gamma_1(GPA) + \gamma_2(year) + \gamma_3(term) + \epsilon \tag{2}$$

AL performance is Average Test Scores. Overall course performance is also used as a robustness check.

X is, in separate regressions, either **Assignment grades** (video-lecture completion and grades in reading, learning checks, and required practice tests, summarized on Table 7) or a combination of **Time on task** (average minutes reading or watching, hours of LMS activity, or LMS page views) and **Timeliness** (percent of late assignments), summarized on Table 7. To prevent averages from being dominated by the zeros of missed assignments, only students who completed the course are included.

4.3.2. Results

Table 8 reports that Practice Tests and the number of Practice Attempts are always predictive of Test scores. Reading scores and video completion predict above-median students' Test scores. These results suggest study strategies for each kind of student: the better students should focus on reading better and watching the whole video, while weaker students should aim for more than the minimum practice score.

Table 9 reports that more time watching the videos and doing the reading predicts better scores; time on the learning management system or lateness are not significant. This suggests that the grade improvements are not just about more work but about the quality of the work. Above-median students should re-watch videos; below-median students should spend more time reading.

In principle, scores in assignments could be highly correlated. Indeed, controlling for video completion and reading scores, homework and practice scores vary little for A students. Interestingly, this correlation breaks down among students predicted to fail the tests. Controlling for video and reading performance, those who scored above-average in the homework passed the class, while those

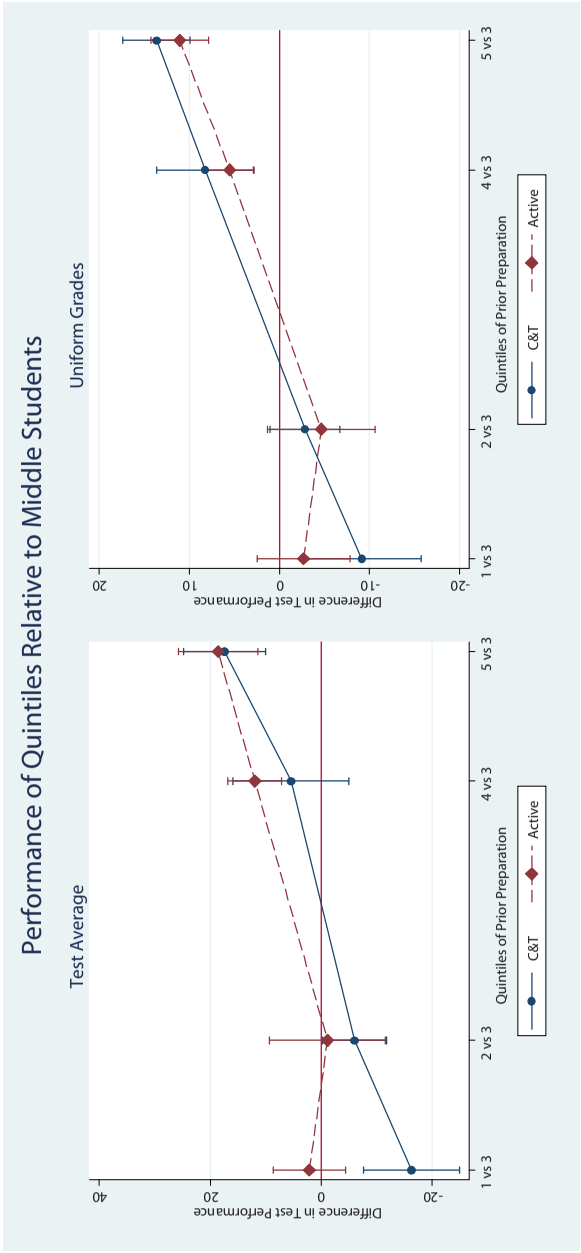

Fig. 7. Average effects of AL course style for quintiles of course.

Table 7
Summary statistics for assignment scores and for time on task and timeliness in Active courses, above-median and below-median GPA students.

Above Median GPA	**Mean**	**SD**	**Min**	**p25**	**p50**	**p75**	**Max**
Video Completion %	81.4	27.4	8.7	69.2	95.2	99.9	100
Reading scores	90.6	14.1	44	88.1	96.7	99.6	100
Check scores	94.1	3.8	85.5	91.4	95.3	97.3	100
Practice Scores	78.6	4.2	71.4	75.6	78.4	81	91.7
Test Attempts	3	2	1	1.7	2.5	3.7	11.7
Avg Video Time	11.2	4.3	0.5	9	11.8	14.4	18.9
Median Reading Time	47.1	34.8	7	23	33.5	66.5	139.5
Hours on Canvas	73.6	30.1	29	56.9	68.9	85.2	186.9
Canvas Page Views	1322.2	696.9	419	891	1135	1738	4187
% Late Assignments	11.3	10.7	0	2.8	7.5	18.7	44.9
Below Median GPA	**Mean**	**SD**	**Min**	**p25**	**p50**	**p75**	**Max**
Video Completion %	60.7	33.1	9	27.5	66.2	92.2	100
Reading scores	69.1	23.5	6.4	56.9	75.1	88.1	98.6
Check scores	89.3	3.9	82.8	85.5	89.2	92.5	97.2
Practice Scores	75.7	5.1	56.4	74.4	76.2	79	84.9
Test Attempts	3.6	1.7	1	2.3	3	5	7.7
Avg Video Time	7.9	4.8	0.8	3.5	7.9	12	15.3
Median Reading Time	22.8	14	0	13.5	21	33	54.5
Hours on Canvas	62	25	25.5	46.6	57.3	69.3	168.5
Canvas Page Views	1256.9	468.1	582	923	1225	1569	2934
% Late Assignments	33.2	24.1	0.9	12.1	31.8	51.4	79.4

who scored much below-average in homework failed the class. This result highlights the make-or-break effect of the Active Learning system on bottom students.

5. Conclusions

A system that enforced spacing, self-testing, elaborative questioning, and proper foundation helped many students improve their performance. Top students solidified their A and many middling and low-end students improved their grades by two-thirds of a letter grade. On the other hand, failure or withdrawal became more common, especially at the low-end of the predicted-grade distribution, possibly due to students accustomed to a coast-and-cram study strategy.

When the strategy was applied partially (e.g., no videos) in other courses, students requested that the missing components be reintroduced, suggesting that the system (rather than any particular component) is responsible for the results. The educational technologies used (Panopto, Perusall, Canvas) are high-quality, user-friendly, and effective. While some components require a high time commitment (especially the videos), Perusall auto-scoring and test banks reduce the time investment. Future research could improve the degree to which students learn from self-testing (e.g., encouraging student reflection on self-test results) or could explore application of other principles from Brown et al. (2014), such as interleaving and attempting solutions before explanation.

Beyond the final grade in the semester, there are three possible benefits of the system. One is long-term learning: more spacing and more self-testing aid long-term retention and transfer of course knowledge to applications (cf. Chew and Cerbin, 2021, 12), which could be tested with grades in further courses, especially in intermediate microeconomics. A second possible benefit is improved work habits taught by an enforced routine.

A final possible benefit is greater satisfaction from more learning and increased confidence before tests. In spite of the frustrations involved in the system or because of increased confidence, course evaluation results stayed at a 4.7 on a 5-point scale, against a 4.3 average for the university. In the words of one of the students,

> We … were taken through in a step by step process throughout the course. In class, the lessons would be gone over in a detailed and easy to understand way. The homework allowed us to think and take in the material given to us in a way that was not overwhelming. We were given multiple tries at the homework so the if we were struggling, we would have a chance to fix ourselves. Everything about this course was aimed for our success!.

CRediT authorship contribution statement

Gabriel Martinez was responsible for Conceptualization, Data curation, Formal analysis, Investigation, Methodology, administration of the project, Resources, Visualization, and all writing. The statistical software used is Stata®. The bibliographic software is Endnote®. The learning management system is Canvas®. The video production and hosting system is Panopto®. The readings were accessible through Perusall®. These were all obtained by Martinez's employer for his use in teaching and research. No funding was needed. No additional supervision or validation was needed or carried out.

Table 8
How assignments contribute to Test scores in Active Courses, controlling for GPA, year, and term. Standard errors in parentheses, $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Below	Below	Below	Below	Above	Above	Above	Above	All	All	All	All
Video Compl.	0.000352	0.00489	0.0780		0.111**	0.107**	0.102**		0.0761+	0.0783*	0.0857*	
	(0.0742)	(0.0711)	(0.0668)		(0.0336)	(0.0375)	(0.0300)		(0.0393)	(0.0390)	(0.0331)	
Reading	0.0182		-0.0501		0.226**		0.181**		0.0798		0.0152	
	(0.0886)		(0.0796)		(0.0676)		(0.0623)		(0.0570)		(0.0491)	
Checks		0.0263		-0.413		0.555+		0.595*		0.454		0.130
		(0.510)		(0.447)		(0.287)		(0.272)		(0.295)		(0.266)
Practice Test			0.705*	0.643*			0.519*	0.518*			0.716***	0.689***
			(0.316)	(0.299)			(0.196)	(0.236)			(0.191)	(0.198)
Test Attempts			-2.778**	-2.739**			-0.969*	-1.438**			-2.069***	-2.064***
			(0.924)	(0.918)			(0.430)	(0.490)			(0.492)	(0.515)
GPA	-0.257	-0.207	1.905	2.905	15.46**	18.62***	12.78**	17.69***	6.048**	5.536*	4.675*	6.056**
	(4.393)	(4.477)	(3.788)	(3.839)	(4.609)	(4.871)	(4.292)	(4.849)	(2.119)	(2.236)	(1.797)	(1.886)
N	39	39	39	39	39	39	38	38	78	78	77	77
Adj. R^2	-0.0767	-0.0779	0.218	0.228	0.669	0.603	0.751	0.644	0.318	0.322	0.528	0.489

Table 9
How time on task and timeliness contribute to Test scores in Active Courses, controlling for GPA, year, and term. Standard errors in parentheses, $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Below	Below	Below	Below	Above	Above	Above	Above	All	All	All	All
Vid. Compl.	1.701*	2.303**	0.533	1.507+	0.682**	0.715**	0.782**	0.782**	1.123*	1.409**	0.592+	1.131*
	(0.798)	(0.754)	(0.628)	(0.793)	(0.242)	(0.245)	(0.255)	(0.255)	(0.442)	(0.419)	(0.333)	(0.459)
Read Time	0.628*		0.448*	0.740**	0.0819*		0.0674+	0.0674+	0.169+		0.169**	0.168+
	(0.248)		(0.185)	(0.253)	(0.0355)		(0.0373)	(0.0373)	(0.0851)		(0.0616)	(0.0877)
LMS Time		0.147				0.0112				0.117+		
		(0.124)				(0.0384)				(0.0678)		
Page Views			-0.00537				0.00185	0.00185			-0.00154	
			(0.00498)				(0.00156)	(0.00156)			(0.00226)	
% Late				0.259								-0.00780
				(0.165)								(0.113)
N	44	46	41	44	36	39	36	36	80	82	77	80
Adj. R^2	0.289	0.303	0.132	0.315	0.564	0.545	0.570	0.570	0.367	0.418	0.369	0.359

Declaration of Competing Interest

No actual or potential conflict of interest was reported by the authors.

Data availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available. The university's Institutional Review Board classified this study as Exempt from the requirements of the Basic HHS Policy for Protection of Human Research Subjects, per 45 CFR §46.104, under exemption criterion 1: educational research. The IRB reference number assigned to this project is IN-011722.

References

Allgood, Sam, McGoldrick, KimMarie, 2020. How can economists use the cognitive challenges framework to enhance economic education? J. Econ. Educ. 1–12.

Anderson, Gordon, Benjamin, Dwayne, Fuss, Melvyn A., 1994. The determinants of success in university introductory economics courses. J. Econ. Educ. 25 (2), 99–119. https://doi.org/10.1080/00220485.1994.10844820.

Asarta, Carlos J., Chambers, Rebecca G., Harter, Cynthia, 2021. Teaching methods in undergraduate introductory economics courses: results from a sixth national quinquennial survey. Am. Econ. 66 (1), 18–28.

Asarta, Carlos J., Schmidt, James R., 2013. Access patterns of online materials in a blended course. Decis. Sci. J. Innov. Educ. 11 (1), 107–123.

Balaban, Rita A., Gilleskie, Donna B., Tran, Uyen, 2016. A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. J. Econ. Educ. 47 (4), 269–287. https://doi.org/10.1080/00220485.2016.1213679.

Becker, William E., Watts, Michael, 1996. Chalk and talk: a national survey on teaching undergraduate economics. Am. Econ. Rev. 86 (2), 448–453.

Becker, William E., Watts, Michael, 2001. Teaching economics at the start of the 21st century: still chalk-and-talk. Am. Econ. Rev. 91 (2), 446–451.

Berrett, Dan, 2015. The unwritten rules of college. Chron. High. Educ. 62 (4), A26–A29.

Boyle, Austin, Goffe, William L., 2018. Beyond the flipped class: the impact of research-based teaching methods in a macroeconomics principles class. AEA Pap. Proc. 108, 297–301. https://doi.org/10.1257/pandp.20181052.

Brown, Peter C., Roediger, Henry L., McDaniel, Mark A., 2014. Make it Stick. Harvard University Press.

Calimeris, Lauren, Kosack, Edward, 2020. Immediate feedback assessment technique (IF-AT) quizzes and student performance in microeconomic principles courses. J. Econ. Educ. 51 (3–4), 211–226.

Calimeris, Lauren, Sauer, Katherine M., 2015. Flipping out about the flip: all hype or is there hope? Int. Rev. Econ. Educ. 20, 13–28. https://doi.org/10.1016/j.iree.2015.08.001.

Caviglia-Harris, Jill, 2016. Flipping the undergraduate economics classroom: using online videos to enhance teaching and learning. South. Econ. J. 83 (1), 321–331. https://doi.org/10.1002/soej.12128.

Cepeda, Nicholas J., Pashler, Harold, Vul, Edward, Wixted, John T., Rohrer, Doug, 2006. Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychol. Bull. 132 (3), 354–380.

Chew, Stephen L., Cerbin, William J., 2021. The cognitive challenges of effective teaching. J. Econ. Educ. 52 (1), 17–40. https://doi.org/10.1080/00220485.2020.1845266.

Cosgrove, Sarah B., Olitsky, Neal H., 2020. Research-based instructional strategies in a flipped principles of microeconomics classroom. Int. Rev. Econ. Educ. 33, 100175 https://doi.org/10.1016/j.iree.2020.100175.

Craft, Erik, Linask, Maia, 2020. Learning effects of the flipped classroom in a principles of microeconomics course. J. Econ. Educ. 51 (1), 1–18. https://doi.org/10.1080/00220485.2019.1687372.

Doyle, Terry, 2008. Helping Students Learn in A Learner-centered Environment: a Guide to Facilitating Learning in Higher Education. Stylus Publishing LLC,.

Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., Willingham, D.T., 2013. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. Psychol. Sci. Public Interest.: A J. Am. Psychol. Soc. 14 (1), 4–58, 10.1177/1529100612453266.

Dunning, David, 2011. The Dunning–Kruger effect: on being ignorant of one's own ignorance. In: Advances in Experimental Social Psychology. Elsevier, pp. 247–296.

Emerson, Tisha L.N., English, Linda K., McGoldrick, KimMarie, 2018. The high costs of large enrollment classes: can cooperative learning help? East. Econ. J. 44 (3), 455–474.

Fiorella, L., Mayer, R.E., 2015. Learning as a Generative Activity: Eight Learning Strategies That Promote Understanding. Cambridge University Press.

Fiorella, Logan, Mayer, Richard E., 2016. Eight ways to promote generative learning. Educ. Psychol. Rev. 28 (4), 717–741.

Fisher, Robert, 1998. Thinking about thinking: developing metacognition in children. Early Child Dev. Care 141 (1), 1–15.

Galizzi, Monica, 2010. An assessment of the impact of online quizzes and textbook resources on students' learning. Int. Rev. Econ. Educ. 9 (1), 31–43.

Gholami, Valeh, Moghaddam, Mostafa Morady, 2013. The effect of weekly quizzes on students' final achievement score. Int. J. Mod. Educ. Comput. Sci. 5 (1), 36.

Goffe, William L., 2021. Online implementation of portions of "the cognitive challenges of effective teaching". J. Econ. Educ. 52 (1), 82–88.

Grimes, Paul W., 2002. The overconfident principles of economics student: an examination of a metacognitive skill. J. Econ. Educ. 33 (1), 15–30.

Guskey, Thomas, 2015. Mastery learning. In: Smelser, Neil J., Baltes, Paul B. (Eds.), International Encyclopedia of the Social & Behavioral Sciences, pp. 752–759.

Henderson, Amanda, Harrison, Penny, Rowe, Jennifer, Edwards, Sam, Barnes, Margaret, Henderson, Simon, 2018. Students take the lead for learning in practice: a process for building self-efficacy into undergraduate nursing education. Nurse Educ. Pract. 31, 14–19.

Hernández-Julián, Rey, Peters, Christina, 2012. Does the medium matter? Online versus paper coursework. South. Econ. J. 78 (4), 1333–1345.

Hossain, Belayet, Tsigaris, Panagiotis, 2015. Are grade expectations rational? A classroom experiment. Educ. Econ. 23 (2), 199–212.

Josephson, Anna, DeBoer, Larry, Nelson, Dave, Zissimopoulos, Angelika, 2019. Learning to do: facilitating practice in a large introductory macroeconomics class. J. Econ. Educ. 50 (2), 142–156.

Karpicke, Jeffrey D., Butler, Andrew C., Roediger III, Henry L., 2009. Metacognitive strategies in student learning: do students practise retrieval when they study on their own? Memory 17 (4), 471–479.

Kruger, Justin, Dunning, David, 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J. Personal. Soc. Psychol. 77 (6), 1121.

Latif, Ehsan, Miles, Stan, 2020. The impact of assignments and quizzes on exam grades: A difference-in-difference approach. J. Stat. Educ. 28 (3), 289–294.

Maclean, Gillis, McKeown, Paul, 2013. Comparing online quizzes and take-home assignments as formative assessments in a 100-level economics course. N. Z. Econ. Pap. 47 (3), 245–256.

McGoldrick, KimMarie, Schuhmann, Peter W., 2016. The impact of challenge quizzes on student knowledge. Am. Econ. Rev. 106 (5), 373–377.

Michinov, Nicolas, Brunot, Sophie, Bohec, Olivier Le, Juhel, Jacques, Delaval, Marine, 2011. Procrastination, participation, and performance in online learning environments. Comput. Educ. 56 (1), 243–252.

Millea, Meghan, Grimes, Paul W., 2002. Grade expectations and student evaluation of teaching. Coll. Stud. J. 36 (4), 582–591.

Miller, Laurie A., Asarta, Carlos J., Schmidt, James R., 2019. Completion deadlines, adaptive learning assignments, and student performance. J. Educ. Bus. 94 (3), 185–194.

Miller, Laurie A., Schmidt, James R., 2021. The effects of online assignments and weekly deadlines on student outcomes in a macroeconomics course. Am. Econ. 66 (1), 46–60.

Nowell, Clifford, Alston, Richard M., 2007. I thought I got an A! Overconfidence across the economics curriculum. J. Econ. Educ. 38 (2), 131–142.
Olitsky, Neal H., Cosgrove, Sarah B., 2016. The better blend? Flipping the principles of microeconomics classroom. Int. Rev. Econ. Educ. 21, 1–11. https://doi.org/10.1016/j.iree.2015.10.004.
Picault, Julien. 2019. The Economics Instructor's Toolbox. https://www.theecontoolbox.com/flipped-classroom. (Accessed 4 March).
Pyc, M.A., Agarwal, P.K., Roediger, H.L., 2014. Test enhanced learning. In: Benassi, V.A., Overson, C.E., Hakala, C.M. (Eds.), Applying Science of Learning in Education: Infusing Psychological Science Into the Curriculum. Society for the Teaching of Psychology, Washington, DC, pp. 78–90.
Roediger, Henry L., Karpicke, Jeffrey D., 2006. The power of testing memory: basic research and implications for educational practice. Perspect. Psychol. Sci. 1 (3), 181–210.
Roediger, Henry L., Pyc, Mary A., 2012. Inexpensive techniques to improve education: applying cognitive psychology to enhance educational practice. J. Appl. Res. Mem. Cogn. 1 (4), 242–248.
Sauvé, Louise, Fortin, Anne, Viger, Chantal, Landry, France, 2018. Ineffective learning strategies: a significant barrier to post-secondary perseverance. J. Furth. High. Educ. 42 (2), 205–222.
Sawler, James, 2021. Economics 101-ism and the Dunning-Kruger effect: reducing overconfidence among introductory macroeconomics students. Int. Rev. Econ. Educ. 36, 100208.
Schlösser, Thomas, Dunning, David, Johnson, Kerri L., Kruger, Justin, 2013. How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. J. Econ. Psychol. 39, 85–100.
Schwartz, Daniel L., Tsang, Jessica M., Blair, Kristen P., 2016. The ABCs of How We Learn: 26 Scientifically Proven Approaches, How They Work, and When to Use Them. WW Norton & Company.
Sheridan, Brandon J., Smith, Ben, 2020. How often does active learning actually occur? Perception versus reality. AEA Pap. Proc.
Tait, Hillary, Entwistle, Noel, 1996. Identifying students at risk through ineffective study strategies. High. Educ. 31 (1), 97–116.
Watts, Michael, Becker, William E., 2008. A little more than chalk and talk: results from a third national survey of teaching methods in undergraduate economics courses. J. Econ. Educ. 39 (3), 273–286.
Williamson, E.G., 1937. To avoid waste. J. High. Educ. 8 (2), 64–70.
Williamson, E.G. 1939. Ineffective Study Habits.
Winkelmes, Mary-Ann, Bernacki, Matthew, Butler, Jeffrey, Zochowski, Michelle, Golanics, Jennifer, Harriss Weavil, Kathryn, 2016. A teaching intervention that increases underserved college students' success. Peer Rev. 18 (1/2), 31–36.
Wozny, Nathan, Balser, Cary, Ives, Drew, 2018. Evaluating the flipped classroom: a randomized controlled trial. J. Econ. Educ. 49 (2), 115–129.
Yamarik, Steven, 2007. Does cooperative learning improve student learning outcomes? J. Econ. Educ. 38 (3), 259–277.
Yandell, Dirk, S., 2017. Grade expectations and overconfidence: is economics different? J. Econ. Educ. 17 (2), 18–28.