

Beyond the Flipped Class: the Impact of Research-Based Teaching Methods in a  
Macroeconomics Principles Class

Austin Boyle  
Penn State University  
aboyle@psu.edu

William L. Goffe  
Penn State University  
bill.goffe@psu.edu

December, 2017

You will find two versions below. The next 5 pages are the penultimate version that is scheduled to appear in the AEA Papers and Proceedings. Following that is a 25 page paper with complete details.

# Beyond the Flipped Class: the Impact of Research-Based Teaching Methods in a Macroeconomics Principles Class

By AUSTIN BOYLE AND WILLIAM L. GOFFE\*

A very active group of STEM education researchers have developed and deployed numerous teaching innovations in the last few decades, significantly changing STEM instruction [Simkins and Maier, 2008]. Some have shown learning gains on the order of 2 standard deviations [Hake, 1998, Deslauriers et al., 2011]. These innovations, often dubbed “research-based” or “evidence-based” teaching methods, employ findings from cognitive science. Thus, their underpinnings are more fundamental than “flipping” or “active learning.” Might employing these findings in an economics classroom lead to similar gains? This paper addresses this question in a redesigned macro principles course with hundreds of students taught by the second author.

## I. Limited Learning by Economics and Physics Students

Walstad and Allgood [1999] show that students are not retaining much knowledge from their economics courses. A similar finding is suggested by the norming process of the “Test of Understanding of College Economics,” (TUCE) [Walstad et al., 2007]. Thousands of students from dozens of institutions show modest improvement on this 30-question assessment over the course of the semester. On the micro version, students went from answering 9.39 questions correctly at the start of the semester to 12.77 at the end, while macro students went from 9.80 to 14.19.

\* Boyle: Penn State University, 307 Kern Graduate Building, University Park, PA 16802, aboyle@psu.edu. Goffe: Penn State University, 304 Kern Graduate Building, University Park, PA 16802, bill.goffe@psu.edu. Thanks to William B. Walstad for access to the program used in Walstad and Wagner [2016], Tsan-Kuang Lee and Rahul Patnaik for research assistance, and Scott Simkins and Mark Maier for frequent discussions. Ben Smith and Carl Wieman provided valuable suggestions on an earlier draft. All remaining errors are ours.

Measured by citations, Walstad and Allgood [1999] has had a modest impact: as of December, 2017, Google Scholar reports 113 while a comparable paper in physics, Hestenes et al. [1992], has 3,370 citations. One potential reason for this paper’s impact is physicists were deeply troubled by students’ difficulty with the “Force Concept Inventory” (FCI), introduced in this paper. This specialized type of assessment, a “concept inventory” contrasts expert thinking with known novice views on everyday phenomena with conceptual questions<sup>1</sup>. Summarizing FCI scores from a large sample of students the authors state,

Since the students have evidently not learned the most basic Newtonian concepts, they must have failed to comprehend most of the material in the course. They have been forced to cope with the subject by rote memorization of isolated fragments and by carrying out meaningless tasks... This gloomy assessment is not intended as a wholesale indictment of the many dedicated and competent physics teachers. *It does tell us, though, that effective instruction requires more than dedication and subject knowledge. It requires technical knowledge about how students think and learn. [emphasis added]*

Another influential STEM paper is Hake [1998]<sup>2</sup>. It found that students taught with a specific type of active learning, “interactive engagement” showed much larger improvements on the FCI and its predecessor than students who were taught with

<sup>1</sup>Physicists have developed more than 80 additional concept inventories.

<sup>2</sup>It has 4,631 Google citations as of October 2017.

straight lecture; the effect size was about 1.8 standard deviations. The principle of “constructivism” from cognitive science explains this result, while “deliberate practice” explains the 2.5 standard deviation gain in a third influential paper, Deslauriers et al. [2011]. These principles and others are described in the next section.

## II. Key Cognitive Science Principles

“Constructivism” includes the idea that students come to the classroom with pre-conceptions and incorrect ones can be difficult to dislodge. Hake [1998] illustrates that just telling students about their misconceptions does not lead to their displacement (the FCI was designed to pick up misconceptions). “Deliberate practice” is how experts become so—with challenging tasks on which they receive timely and accurate feedback that help them enrich their understanding. Its classroom application in Deslauriers et al. [2011] leads to large learning gains. “Schemas” are how humans store knowledge; experts have much denser connections between concepts than novices. “Working memory” are the ideas and procedures that can be held in short-term memory; it is much more limited in novices than in experts. A class based solely on lecture will saturate the working memory of a novice. “Retrieval practice” is how memory is reinforced by recalling a fact or procedure. “Interleaving” is studying dissimilar material together while “spacing” focuses on similar topics with time gaps between sessions. “Metacognition” is thinking about one’s own thinking—experts do it as a matter of course while the concept can be foreign to novices. “Desirable difficulty” is the idea that the mind stores concepts that took some effort to understand. Professors suffer from the “curse of knowledge,” which is when an expert literally cannot think like a novice given their different schema and larger working memory. Brown et al. [2014], Ericsson and Pool [2016], Moulton [2014] detail these principles. They were incorporated into the class redesign described next.

## III. Redesigned Class Structure

### A. Student Preparation for Class: JiTTs

Instruction on a topic began with students reading ahead and answering essay questions online with “JiTt” (Just in Time Teaching) assignments [Simkins and Maier, 2009]. JiTTs were assigned every two or three weeks as the class approached a new topic. Following Mazur and Watkins [2009], the last question in a JiTT was typically “What did you find confusing, interesting, or surprising in this reading?” This encouraged student metacognition and the answers alert the instructor to student difficulties that they might not know about due to the curse of knowledge.

### B. Clicker Questions and Worksheets

About 5 clicker questions were asked per 50-minute class totaling 200 questions for the semester. Wieman et al. [2017] list the most effective types of clicker questions as: “reveal pre-existing thinking,” “test conceptual understanding,” “apply ideas in new context/explore implications,” “predict results of lecture demo,” “draw on knowledge from everyday life,” and “relate different representations (graphical, mathematical, ...).” The vast majority of questions were one of these types.

Questions were challenging—about half of student responses were correct<sup>3</sup>. These desirable difficulties help students retain question content. The questions also invoke deliberate practice, retrieval practice, schema generation (when connections are made between topics) and potentially address prior knowledge (part of constructivism).

“Peer Instruction,” was used when 30–70% of students answered questions correctly. The initial vote summary was hidden from students and the instructor requested that students try to convince their peers of the correct answer with their reasoning. Students voted again after a minute

<sup>3</sup>Students were engaged without grading these questions for correctness.

or two. Almost always, the class moved toward the correct answer (the mean improvement was 20%). Following the revote, the instructor thoroughly described the correct answer and the reasoning behind it.

Since many of the clicker questions were conceptual,<sup>4</sup> students received little in-class feedback on performing actual calculations or drawing graphs. To address this issue worksheets were used over the semester.

### C. Lecture

While considerable time in class was devoted to clicker questions on topics that students read outside of class, there was still substantial lecture. One particular relevant finding from cognitive science to lecture are the constraints of human working memory. Moulton [2014] puts it this way: “Our minds have severe, inherent limits on the amount of information they can consciously process at any given time and this limit varies dramatically based on our experience with and strategies for processing that information.” To deal with these limitations, an outline for the day was projected on an auxiliary screen throughout the class and it was frequently referred to by the instructor. PowerPoint slides were structured so that content came in line by line or curve by curve. Pains were made to connect new ideas to previous ones and thus connect to pre-existing ideas. Where possible, concrete examples preceded abstract ideas [Willingham, 2009, Ch. 4].

### D. Quizzes and Exams

Rather than the traditional two or three midterms, seven quizzes were given every 2 weeks with exam-caliber questions. As each quiz only took a portion of the class time, correct answers were revealed to students after quizzes were submitted. After consulting with their peers, students voted on the questions they wished to see explained in more detail. The combination of feedback and thinking about what they did not

<sup>4</sup>One example was “Say that inflation fell. Then the price level (measured by the CPI or GDP deflator) would be sure to fall as well.”

understand, after consultation with their peers, constitute deliberate practice and metacognition. Also, after the first quiz, each succeeding one contained questions on topics from previous quizzes. Thus, spacing was employed and a comprehensive final exam allowed for one more round of it.

### E. Quiz Reflections

About a week after each quiz a brief “quiz reflection” (also known as an “exam wrapper”) was assigned. To aid student metacognition, students were asked brief questions of two types. 1) Which quiz question they found most difficult on the previous quiz, if they now understood it, and what they would do if they did not fully understand it. The selected topics frequently surprised the instructor; this would seem to be the curse of knowledge. 2) Students were asked about how they studied and how they might improve their study methods.

### F. Homework

There were five homeworks during the semester; all questions were instructor-written. Topics were shuffled to achieve interleaving and each homework after the first contained questions from previous parts of the course to implement spacing. Wieman [2016] was used extensively: for calculations students had to select the appropriate data from a table for the entire homework, they were asked to make qualitative judgments on an answer being reasonable, and to identify what concepts needed to be used.

## IV. Empirical Results

Learning gains over the semester were compared to gains in the TUCE norming population, comprised of roughly 80<sup>5</sup> self-selected instructors, presumably with considerable interest in teaching well. Keep in mind that their teaching methods and possible student performance incentives on the

<sup>5</sup>Walstad et al. [2007] lists 157 instructors as giving the TUCE during its norming, but keep in mind that there are both microeconomic and macroeconomic versions.

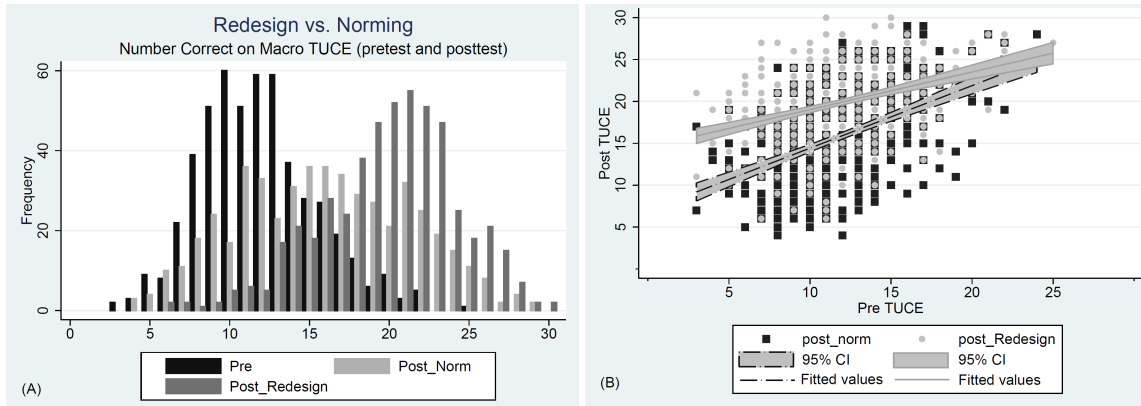


FIGURE 1. MORE IMPROVEMENT WITH ACTIVE LEARNING: (A) SHOWS IMPROVEMENT IN THE REDESIGNED CLASS VS. A SAMPLE FROM THE NORMING POPULATION MIMICKING THE REDESIGNED CLASS’S PRE-TUCE SCORES. (B) SHOWS STATISTICALLY SIGNIFICANT IMPROVEMENTS IN STUDENTS WITH A PRE-SCORE OF 19 OR LOWER (97% OF STUDENTS).

TUCE are unknown. In this study the post-TUCE was a bonus attached to the final exam, which could lead to an upward bias in the results.

Two sections of the second author’s principles of macroeconomics were taught with the above methods in Fall 2015. 636 students took the final exam where the TUCE was offered as a bonus. 508 of these students also took the TUCE at the start of the semester.

Thus, there is pre-post TUCE data for 80% of those who completed the course. The post-TUCE results are basically the same for those who took the pre-TUCE and those who did not. The mean pre-TUCE score was 11.90 questions correct (out of 30) and the mean post-TUCE score was 19.91. The mean learning gain,  $g = (post - pre)/(30 - pre)$ , was .44. In the TUCE norming population, the mean pre-TUCE score was 9.80 questions correct and the mean post-TUCE score was 14.19, for a gain of .22. Thus, students in this study had *twice* the gain of the norming population.

One notable difference between the norming population and the students in this study is the pre-TUCE mean scores: 9.80 versus 11.90. Perhaps students with a higher pre-TUCE score are capable of learning more.

Thus, data from the norming population

were sampled to create a “norming sample” mimicking pre-TUCE scores from the redesigned course described above. Figure 1 shows the distributions of pre-TUCE scores and post-TUCE scores from both the redesigned class and the norming sample. The difference in the means of post-TUCE between the redesigned course and the norming sample is 3.93, representing .82 standard deviations. The 99.5% confidence interval for this difference is from 3.20 to 4.66.

Finally, Figure 1 compares the matched distribution from the norming sample to students in the redesigned course. Note how for a given pre-TUCE score, students in the redesigned course generally had higher post-TUCE scores.

## V. Conclusion

The empirical results in this study suggest that using a research-based approach to teaching principles of macroeconomics might improve learning. The results presented were both statistically and empirically significant with an effect size of .82 standard deviations. Besides promising empirical gains, research-based teaching methods provide instructors considerable guidance on teaching and learning. Writing JiTT and clicker questions and viewing the resulting answers forces the instructor to reflect upon challenging topics. The curse of

knowledge is addressed multiple times per week of instruction.

Numerous teaching methods that seem to be fairly uncommon in economics classrooms were implemented with promising results. Given the experimental design, it is not possible to say which ones matter most. However, three innovations might be the most important as they seem to be uncommon in economics classrooms. First, students received considerable feedback via clicker questions and quizzes compared to a lecture-oriented class. Second, there was considerable spacing of learning—many topics were tested several times over the semester. Third, many of the clicker questions were conceptual rather than computational so they were higher on Bloom's Taxonomy. Fourth, students reflected on their quizzes, which should aid metacognition. Finally, research opportunities abound for economic education researchers interested in further investigating these methods.

#### REFERENCES

- William B. Walstad and Jamie Wagner. The disaggregation of value-added test scores to assess learning outcomes in economics courses. *The Journal of Economic Education*, 47(2):121–131, 2016.
- Scott P. Simkins and Mark H. Maier. Learning from physics education research: Lessons for economics education. *Economics Research Network*, 2008.
- Richard R. Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66:64–74, 1998.
- Louis Deslauriers, Ellen Schelew, and Carl Wieman. Improved learning in a large-enrollment physics class. *Science*, 332(6031):862–4, 2011.
- William B. Walstad and Sam Allgood. What do college seniors know about economics? *The American Economic Review*, 89(2):350–354, May 1999.
- William B. Walstad, Michael W. Watts, and Ken Rebeck. *Test of Understanding in College Economics: Examiner's Manual*. National Council On Economic Education, fourth edition, 2007.
- David Hestenes, Malcolm Wells, and Gregg Swackhamer. Force concept inventory. *The Physics Teacher*, 30(3):141–158, 1992.
- Peter C. Brown, Henry L. Roediger III, and Mark A. McDaniel. *Make It Stick: The Science of Successful Learning*. Harvard University Press, 2014.
- Anders Ericsson and Robert Pool. *Peak: Secrets from the new science of expertise*. Houghton Mifflin Harcourt, 2016.
- Samuel T. Moulton. *Applying Psychological Science to Higher Education: Key Findings and Open Questions*. The Harvard Initiative for Learning and Teaching, April 2014.
- Scott P. Simkins and Mark H. Maier. *Just in Time Teaching: Across the Disciplines, and Across the Academy*. Stylus Publishing, 2009.
- Eric Mazur and Jessica Watkins. Just-in-time teaching and peer instruction. In Scott P. Simkins and Mark H. Maier, editors, *Just in Time Teaching: Across the Disciplines, and Across the Academy*. Stylus Publishing, 2009.
- Carl Wieman et al. Clicker resource guide: An instructor's guide to the effective use of personal response systems (clickers) in teaching. *Carl Wieman Science Education Initiative, University of British Columbia*, 2017.
- Daniel T. Willingham. *Why Don't Students Like School?: A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom*. John Wiley & Sons, 2009.
- Carl Wieman. Creating good homework problems (and grading them), September 2016.

# Beyond the Flipped Class: the Impact of Research-Based Teaching Methods in a Macroeconomics Principles Class\*

Austin Boyle  
Penn State University  
aboyle@psu.edu

William L. Goffe  
Penn State University  
bill.goffe@psu.edu

December, 2017

## Abstract

Cognitive scientists have identified numerous fundamental principles that influence learning; these include deliberate practice, interleaving, retrieval practice, spacing, metacognition, desirable difficulties, limited working memory, the curse of knowledge, schema generation, and constructivism. STEM (science, technology, engineering, and mathematics) education researchers have repeatedly shown improved learning when instruction employs these principles. In particular, teaching methods like flipping and clickers work best when implemented using them. These “research-based” (sometimes called “evidence-based”) teaching methods are becoming the norm in STEM teaching. A macro principles course at Penn State was redesigned with these principles in mind. 508 students in this course achieved .77 standard deviations more learning than principles students normally do on the macroeconomic Test of Understanding of College Economics (TUCE).

## 1 Introduction

In the last few decades, a very active group of STEM education researchers have developed and deployed numerous teaching innovations.<sup>1</sup> Some have shown learning gains on the order of 2 standard deviations ([Hake, 1998](#); [Deslauriers et al., 2011](#)). These and related papers have had considerable influence on how STEM instructors view teaching and indeed STEM teaching methods are rapidly changing. Upon hearing of such successful interventions, an economic educator is likely to wonder if these methods might translate into an economics classroom and lead to similar gains. This paper addresses this question.

A partial explanation for these startling gains seen in some STEM education research is that they employ findings from cognitive science, the study of how people learn. This research dates back more than a century and while it is an extraordinarily rich literature<sup>2</sup> with many remarkable and even surprising findings, only now are the results becoming easily accessible. Recent years have seen the publication of books by leading researchers for the public in general and teachers, instructors, and professors in particular ([Ambrose et al., 2010](#); [Bransford et al., 2000](#); [Brown et al., 2014](#); [Ericsson and Pool, 2016](#); [Miller, 2014](#); [Willingham, 2009](#)).<sup>3</sup> In writing this paper, these sources lead to new and exciting ways of thinking about teaching and learning economics.

Economic educators know that learning should be active and that classes likely should be flipped, but exactly how? What specific practices might maximize learning and what might safely be avoided? For example, can clickers be used to increase learning, and if so, what type of questions provide the most benefits? The economics education literature offers few suggestions, but the STEM education literature offers many and some will be

---

\*We would like to thank William B. Walstad for access to the program used in [Walstad and Wagner \(2016\)](#), Tsan-Kuang Lee and Rahul Patnaik for research assistance, and Scott Simkins and Mark Maier for frequent discussions. Eric Hudson, Ben Smith and Carl Wieman provided valuable suggestions on an earlier draft. All remaining errors are ours.

<sup>1</sup>[Simkins and Maier \(2008\)](#) is a fine guide for economists.

<sup>2</sup>Much of it is conducted on college students with random controlled trials.

<sup>3</sup>Fortunately, writing as scholars, these authors have included references to the primary literature.

described and used here. Or, how should economics instructors treat ideas that students bring to the classroom? The economics education literatures is virtually silent on this point while the STEM education research literature is vast. Indeed, physicists have developed three dozen “concept inventories” that measure how instruction impacts ideas students bring to class. Using these assessments, these researchers have found that simply telling students about their incorrect preconceptions can have remarkably little impact on learning and other methods must be used instead. Taken as a whole, STEM education researchers bill these teaching methods built upon their findings as “research-based” or “evidence-based.” These terms are fitting given their extensive empirical base and their foundation built on cognitive science.

This paper explores the impact of commonly-used STEM teaching methods in a redesigned principles of macroeconomics class. Translating these STEM methods turned out to be fairly straightforward. Also, these methods actually made teaching more enjoyable as it was possible to watch learning take place. As cognitive science has proved to be a very fruitful base for STEM education researchers, cognitive science findings not (yet) in widespread use in STEM classrooms were implemented in the redesigned course as well. In short, the redesigned course implemented many methods found in the STEM education and cognitive science literatures.

The redesigned course was taught to two sections of principles of macroeconomics at Penn State in the fall of 2015 by the second author. 636 students completed the two sections (2 and 3) and learning was measured by gains over the semester on the Test of Understanding of College Economics (Walstad et al., 2007); it was given at the start and the end of the semester (“pre” and “post”) and data was collected for 508 students. Learning gains over the semester were compared to gains in the TUCE norming sample; it would appear to be a reasonable control as the roughly 80 instructors<sup>4</sup> were self-selected and presumably have a considerable interest in teaching well.

This paper is organized as follows. Section 2 explores limited learning in economics principles classes, as well as in comparable physics classes. This section concludes with a comparison of how the different professions reacted to this unwelcome news. Section 3 describes some elements of STEM education research, with a particular emphasis on Hestenes et al. (1992), Hake (1998), and Deslauriers et al. (2011). Section 4 summarizes key findings from cognitive science for college instructors. Section 5 describes how a macro principles class was redesigned with concepts from the previous two sections in mind, and the following section describes the empirical results. A conclusion ends the paper.

## 2 Limited Learning by Economics and Physics Students

Economic education researchers have studied many topics. However, one that has achieved relatively little direct attention is limited learning by economics students. This limited learning was perhaps best pointed out by Walstad and Allgood (1999).<sup>5</sup> They used two approaches. In the first, Gallup Inc. surveyed 300 college seniors and asked them 15 basic economic questions. If they had not taken an economics course, they answered 48% correctly, while if they had taken an economics course, they answered 62% correctly. In the second, these researchers examined the economics portion of the “Major Field Test in Business.” This exit assessment is commonly given to graduating business majors and, of course, many take several economics courses. The economic portion of this assessment contains basic principles-level questions, yet the mean score was 41%.

Another suggestion that economics students are not retaining much from their courses comes from the norming process of the “Test of Understanding of College Economics,” (TUCE) (Walstad et al., 2007). More than 3,000 students from approximately 50 institutions took this assessment during its construction. There are both microeconomic and macroeconomic versions, each with 30 questions. It was given at both the start of the semester (“pretest”) and at the end of the semester (“posttest”). The gains over the semester were fairly small; on the macroeconomic version students in the norming sample went from answering 9.39 questions correctly to answering 12.77 correctly (3.38 more questions correct) and on the microeconomic version, from 9.80 to 14.19 correct (4.39 more correct). Figures 1 and 2 show these data and can be seen as confirming the suggestion that for many students, the gain on the TUCE assessment is modest.

<sup>4</sup>Walstad et al. (2007) lists 157 instructors as giving the TUCE during its norming, but keep in mind that there are both microeconomic and macroeconomic versions.

<sup>5</sup>A similar argument was made in Goffe (2013) and Bice et al. (2014).



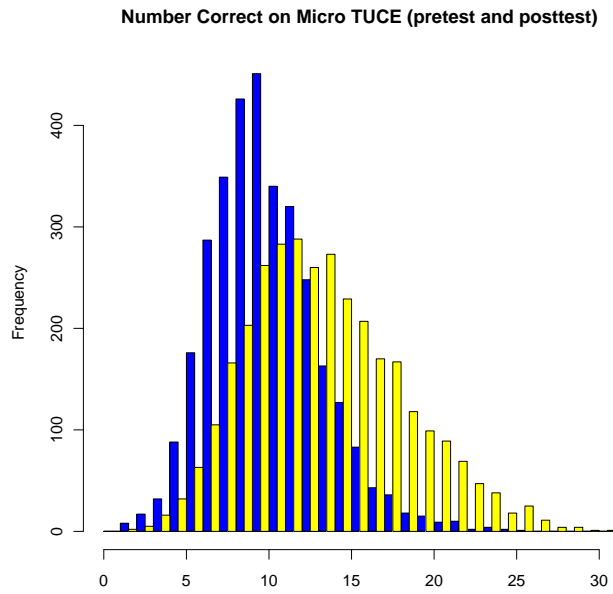


Figure 1: Blue denotes the number correct on the pretest and yellow on the posttest of the TUCE microeconomic norming sample.

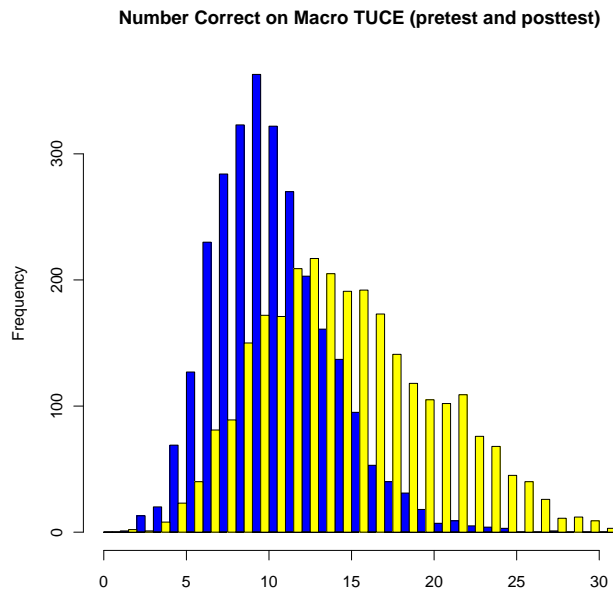


Figure 2: Blue denotes the number correct on the pretest and yellow on the posttest of the TUCE macroeconomic norming sample.

To help set the stage for later sections of this paper, it is useful to see how another discipline, physics, reacted to a paper similar to [Walstad and Allgood \(1999\)](#). [Hestenes et al. \(1992\)](#) also found limited learning but it had a much larger impact. One simplistic measure of impact is, of course, citations. In December 2017, Google Scholar reported 113 citations to [Walstad and Allgood \(1999\)](#) and 3,370 citations to [Hestenes et al. \(1992\)](#). A potential reason for this paper's impact is physicists were deeply troubled by students having difficulty with the "concept inventory" introduced in this paper. This specialized type of assessment contrasts expert thinking with novice views of a topic. Typically, the questions do not involve calculations but instead probe fundamental understanding. The multiple choice "distractors" are typical student views while the correct answers are the views held by experts, or at least students who deeply understand the topic at hand. This paper introduced the "Force Concept Inventory," which measures students' understanding of Newton's Laws, a staple of a first semester physics course. Perhaps the FCI is best illustrated with one of its questions:

- A large truck collides head-on with a small compact car. During the collision:
- (A) the truck exerts a greater amount of force on the car than the car exerts on the truck.
  - (B) the car exerts a greater amount of force on the truck than the truck exerts on the car.
  - (C) neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.
  - (D) the truck exerts a force on the car but the car does not exert a force on the truck.
  - (E) the truck exerts the same amount of force on the car as the car exerts on the truck.

Ideally, concept inventories ask about everyday phenomena about which students might have prior thinking.<sup>6</sup> Summarizing FCI scores from a large sample of students, the authors state,

The implications could not be more serious. Since the students have evidently not learned the most basic Newtonian concepts, they must have failed to comprehend most of the material in the course. They have been forced to cope with the subject by rote memorization of isolated fragments and by carrying out meaningless tasks.

They go on to say,

This gloomy assessment is not intended as a wholesale indictment of the many dedicated and competent physics teachers. It does tell us, though, that effective instruction requires more than dedication and subject knowledge. It requires technical knowledge about how students think and learn.

Six years later, another influential paper, [Hake \(1998\)](#)<sup>7</sup> found that students taught with a specific type of active learning billed as "interactive engagement," performed much better on the FCI and its predecessor than students who were taught with lecture. More specifically, this paper presents learning gains over a semester with the FCI and its predecessor from 62 introductory physics classes with 6,542 students. The classes were in high schools, colleges, and universities. Fourteen classes were taught traditionally (i.e., with lecture) and 48 were taught with "interactive engagement": "designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors."

Mean learning gain per class was calculated with  $g = (a_e - a_s)/(100 - a_s)$  where  $a$  is the percentage score on an assessment with  $a_s$  the score at the start of the semester and  $a_e$  is the score at the end of the semester. The value represents the fraction of the possible gain that is achieved over the semester and it will be between 0 and 1. Figure 3 shows the distribution of gains for traditionally taught courses (in red) and interactive engagement courses (in green).<sup>8</sup> The mean gain for the traditionally taught courses was .23 and it was .48 for interactive

<sup>6</sup>Physicists have developed more than 80 additional concept inventories; see [PhysPort: Browse Assessments \(2017\)](#). Additionally, biologists have developed more than 30; see [Concept Inventories/Conceptual Assessments in Biology \(CABs\)](#).

<sup>7</sup>It has 4,631 Google citations as of December 2017.

<sup>8</sup>The summed vertical heights of traditionally taught courses sums to 1, as does the vertical height of the interactive engagement courses, as both show the distribution of gains for courses taught either way.

engagement courses and the gain was 1.8 standard deviations. Note that no traditionally taught course had a gain of more than about .3 and this is lower than almost every interactive engagement course. With further development of teaching methods in the intervening years, mean gains are now on the order of .6 in active learning classrooms with a surprisingly small variation across instructors (Hoellwarth and Moelter, 2011).<sup>9</sup>

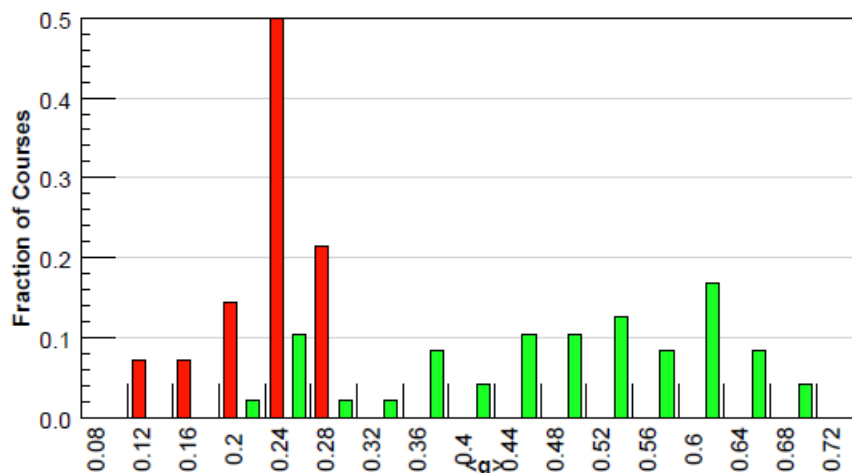


Figure 3: Learning gains by 6,542 physics students in 14 lecture (red/dark grey) and 48 active learning (green/light grey) classes. A higher <g> denotes greater conceptual understanding. Hake (1998).

Hestenes et al. (1992) and Hake (1998) helped instigate a movement away from teaching introductory physics with lecture; in 2008 a survey of 772 physicists who teach these courses Henderson et al. (2012) found that only 28% of respondents had not tried at least one “research-based instructional strategy”<sup>10</sup> and 49% currently use at least one method. Watts and Schaur (2011) conducted a roughly similar survey and it finds that a very large majority of economics classroom time is spent with the instructor lecturing.

### 3 Research-Based Teaching Methods in STEM Classrooms

There are relatively few active learning papers in economics—13 are cited in the 2015 *Journal of Economic Literature Review* review of the economic education literature (Allgood et al., 2015). In contrast, a meta-study of active learning in STEM disciplines (Freeman et al., 2014) uses 225 publications.<sup>11</sup> Further, economists rarely cite cognitive science research on human learning while this is fairly common in the STEM education literature. Succinct summaries of the cognitive science literature as it applies to teaching include Moulton (2014) and Deans for Impact (2015), while book-length introductions include Brown et al. (2014), Willingham (2009), Ambrose et al. (2010), Miller (2014), and Bransford et al. (2000).<sup>12</sup>

One key finding from cognitive science is “constructivism.” Resnick (1983) offers three key criteria for this framework of human learning. First we “construct understanding” of new topics—we do not merely repeat what we are taught. Second, understanding is stored in “schemata,” which are relationships of ideas and facts. Typically, experts have much richer schema than novices. Finally, “learning depends on prior knowledge. Learners try to link new information to what they already know in order to interpret the new material in terms of established

<sup>9</sup>That is, they do not see teacher effects.

<sup>10</sup>These include the methods mentioned above with Hake but also include other non-lecture methods that have since been developed.

<sup>11</sup>They find a positive impact from active learning.

<sup>12</sup>Each of these books counts at least one cognitive scientist among its authors.

schemata.”<sup>13</sup>

The results reported in Hake (1998) nicely illustrate this third point. It uses the FCI and its predecessor to measure student learning of Newton’s Laws, which are highly non-intuitive. Unless instruction explicitly takes into account student’s incorrect prior knowledge, deep learning does not occur for the large majority of students. After reading the physics education research literature on student misconceptions, novice physics instructor Shane Hutson (Hutson, 2014), describes what he found in the classroom:

And it was really apparent that they [students] weren’t blank slates, that they were bringing in a lot of ideas that had some rational basis but weren’t right. And that those ideas that they brought in were really resistant to change, that no matter how many times a lecturer stood in front of the classroom and told them that those ideas weren’t correct, it just didn’t sink in. And so you needed to use one of a variety of active learning techniques in the classroom to bring those misconceptions to the forefront and then be able to address it.<sup>14</sup>

A telling anecdote that illustrates the interplay between lecture, the FCI, and constructivism occurred when Harvard physicist Eric Mazur first gave the FCI to his class in 1990 (Crouch et al., 2007). To this point in his career, he had lectured his very capable students with what he thought were excellent descriptions and he received fine evaluations. A few minutes after he passed out the FCI, a student asked “How should I answer these questions? According to what you taught me or according to the way I usually think about these things?”<sup>15</sup> (Mazur, 2009). While these students could solve sophisticated problems, most did not deeply understand the subject and thus performed poorly on the FCI. Partly as a result of this experience, Mazur has gone on to be a leading physics education researcher.

“Deliberate practice” is another relevant cognitive science concept for educators. It became famous with the “10,000 hour rule” that was popularized in Malcolm Gladwell’s book, *Outliers*. This “rule” somewhat inaccurately portrayed by Gladwell, comes from work by Anders Ericsson and his coauthors, of which the most cited is Ericsson et al. (1993). They find that it takes thousands of hours of practice to reach expert abilities in many disciplines, though the actual number varies. Also largely missing in popular accounts are details of this type of practice. Ericsson labeled it “deliberate practice” and it includes attempting tasks just beyond current abilities, using one’s full attention, timely and accurate feedback, and the development and elaboration of “effective mental representations.” (Ericsson and Pool, 2016, pp. 99-100).

This concept is explicitly invoked in one influential physics education research study, Deslauriers et al. (2011). In it, learning for one week was compared between two sections taking the second semester of introductory physics, electricity and magnetism. Both sections had approximately 270 students and had roughly the same measurable characteristics before the intervention, such as mean midterm and concept inventory scores (in this case, the “Brief Electricity and Magnetism Assessment”) as well as views of scientific reasoning (the “Colorado Learning Attitudes about Science Survey”).

The control section “was lectured by a motivated faculty member with high student evaluations and many years of experience teaching this course” while the experimental section was taught by the first author with assistance from the second. Both of the instructors of the experimental section had never taught their own course but had been teaching assistants. Both also had training in cognitive science as well as physics education research. Students in the experimental section prepared for class by reading brief assignments and then answered true/false questions on the readings in their course management system. During class these students answered challenging clicker questions and they received extensive feedback. Deliberate practice was thus in play.<sup>16</sup>

Remarkably, none of the 14 clicker questions used that week explicitly ask for a calculation. Instead, these questions were of various, often overlapping types: conceptual, testing understanding of different representations of a concept, connecting to prior knowledge, or testing mental models of a concept. These question types are

---

<sup>13</sup>Constructivism plays a key role in “Vision and Change,” a movement among biological societies to incorporate more active learning in undergraduate education (Hartle et al., 2012).

<sup>14</sup>He later adds, “It was a little disconcerting how accurate the literature [on misconceptions] was at first.”

<sup>15</sup>Recall how the distractors in concept inventories like the FCI are common student misconceptions.

<sup>16</sup>Before two of the three classes that week, the novice instructors interviewed students to go over the questions for those days to ensure that they were sufficiently challenging.

consistent with [Resnick \(1983\)](#) who described constructivism above. There she argues that qualitative understanding is essential for deep understanding of scientific problems—robust qualitative problem solving depends upon qualitative understanding. These clicker questions are also similar in spirit to the questions in the FCI in that they tend to probe for deep understanding rather than asking students to solve a problem or calculate an answer.

All instructors agreed on a 12-question assessment to evaluate learning for that week’s material. The score a student earned did not influence their grade, so students had little incentive to study for the assessment. Thus, the assessment presumably measured the classroom component of learning and was not influenced by studying that students carry out on their own. The control, the class that was lectured to, had a mean score of 41%, while the experimental section had a mean score of 74%, for a difference of 2.5 standard deviations. Figure 4 shows the results.

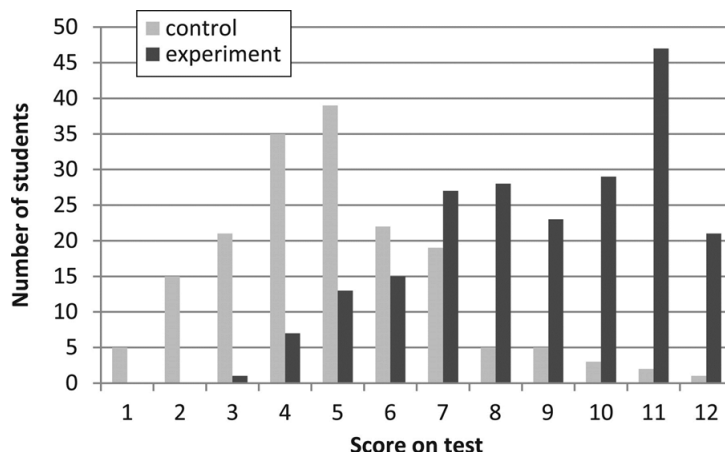


Figure 4: Assessment results between a class taught with lecture (light gray) and one taught with methods based on results from physics education research and cognitive science (dark gray). Deslauriers et al., 2001.

The last author of [Deslauriers et al. \(2011\)](#), and its organizer, Carl Wieman, has taken a career path that would be unusual for an economist. For years he was an experimental physicist and this led to a Nobel Prize in 2001. Well before this award he was also conducting pedagogy research. As he describes in [Wieman \(2007\)](#), he was puzzled why his undergraduate students were not learning much:

When I first taught physics as a young assistant professor, I used the approach that is all too common when someone is called upon to teach something. First I thought very hard about the topic and got it clear in my own mind. Then I explained it to my students so that they would understand it with the same clarity I had. ... At least that was the theory. But I am a devout believer in the experimental method, so I always measure results. And whenever I made any serious attempt to determine what my students were learning, it was clear that this approach just didn't work.

He was even more puzzled over the startling transformation of his graduate students. At the start of their training, “often it seemed that they didn’t even really understand what physics was.” But, in a few years, they were “expert physicists, genuine colleagues.” He goes on to say,

...I tackled it [the rapid transformation of graduate students] like a science problem. I started studying the research on how people learn, particularly how they learn science, to see if it could provide a more satisfactory explanation of the pattern. Sure enough, the research did have another explanation to offer that also solved the earlier puzzle of why my classroom teaching was ineffective.

Deslauriers et al. (2011) illustrates Wieman's solution to the puzzle of why explaining to his students lead to limited learning.

Hake, Mazur, and Wieman are just three of many physics education researchers; Figure 5 shows the nearly 90 U.S. physics departments with "physics education research groups." Perhaps the status of this field is best demonstrated by one of the sister journals of "Physical Review Accelerators and Beams"—"Physical Review Physics Education Research." Both are one of twelve journals of the American Physical Society, the analogue of the American Economic Association.



Figure 5: Physics Departments with Physics Education Research Groups

A particularly useful classroom implementation guide to using these methods with clickers is Wieman et al. (2017). Thoughtful use of these devices is one way to employ deliberate practice as students can try a challenging answer in class and then receive prompt feedback on their thinking. Based on their research, they find that these types of questions are less effective (quoting exactly):

- Quiz on the reading assigned in preparation for the class
- Test recall of lecture point
- Do a calculation or choose next step in a complex calculation
- Survey students to determine background or opinions

while

- Elicit/reveal pre-existing thinking
- Test conceptual understanding
- Apply ideas in new context/explore implications
- Predict results of lecture demo, experiment, or simulation, video, etc.
- Draw on knowledge from everyday life
- Relate different representations (graphical, mathematical, ...)

are more effective.<sup>17</sup>

A later paper by Smith et al. (2011) provides further research-based suggestions on clicker use. They find that learning from clicker questions is maximized when the following occurs. First, students vote on their own. Then, if the results are mixed, students try to convince each other of the correct answer and then revote (i.e., "Peer Instruction," which was popularized by Eric Mazur). Finally, the question and possible answers are discussed by

<sup>17</sup>This categorization of more and less useful question types likely holds for other active learning methods that ask questions.

the instructor. Figure 6 shows the key findings—the vertical axis is a variant of the  $g$  measure used above in [Hake \(1998\)](#) between the original question and a follow-up question on the topic at hand. As one can see, the largest gain in understanding occurs in the “combination mode”—Peer Instruction with an instructor explanation.

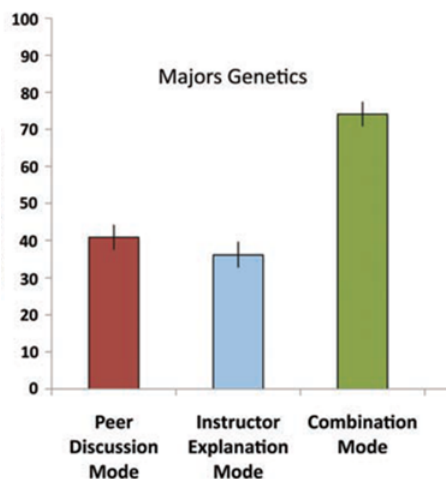


Figure 6: Learning gains with different types of clicker use. Smith et al., 2011.

The methods described above are often labeled “research-based” or “evidence-based” teaching methods. They generally have broad empirical support for specific teaching practices and frequently use findings from cognitive science as their basis. As one can see, they are more nuanced and descriptive than the idea that students should be active or that a class should be flipped.

Rather remarkably, [Henderson et al. \(2015\)](#) suggests that STEM education researchers have yet to fully employ other findings from cognitive science. Nonetheless, the approaches taken by these researchers described above is sufficiently different from how economists typically teach that perhaps economics students might benefit from these methods. Later sections of this paper explore this point.

## 4 Additional Important Teaching Principles from Cognitive Science Research

The previous section described how STEM education researchers have employed constructivism and deliberate practice. However, as [Henderson et al. \(2015\)](#) notes, there are many findings from cognitive science that seem to be rarely deliberately employed or studied in STEM classrooms. As one might guess, cognitive science is a vast subject and indeed there are even undergraduate courses devoted to it. Thus, this list will by necessity be somewhat selective.

One reasonable place to start is [Moulton \(2014\)](#) as the audience is higher education, the paper is succinct, and he cites the primary literature. He argues that the following findings are relevant:

- Retrieval practice (the “testing effect”): being tested on a topic leads to greater retention than restudying that topic.
- Spaced practice: when studying is spread out over time, instead of “massed” (i.e., crammed), long-term retention is enhanced.
- Interleaved practice: mixing up the topics one studies or practices leads to deeper understanding.
- Desirable difficulties: more challenging tasks lead to greater retention of the topic at hand. This and the previous three topics are a major focus of [Brown et al. \(2014\)](#).
- Deep processing: students who study deeper levels of a topic learn more.



- Transfer-appropriate processing: being able to use a learned skill in a different context is fraught with caveats. A student might be able to successfully use a skill in one situation but not in another.
- Deliberate practice: as above.
- Limited capacity and cognitive load theory: it is very easy to overload a student's short-term memory. Experts can process much more than novices given their base of knowledge.
- Transfer-appropriate processing: humans can process verbal and visual information concurrently and this can aid learning.
- The curse of knowledge: it can be very difficult for experts to understand a topic as a novice does. The implications for teaching should be clear.
- Mind wandering: this is particularly common during a continuous lecture.
- Limited self-control: willpower can easily be used up.
- Planning fallacy: humans generally think a given project will take less time than it actually ends up taking.
- Overconfidence: for example, students often overestimate their exam scores. Overconfidence occurs less often when there is frequent feedback.
- Achievement motivation: intrinsically motivated humans learn more than those who have a performance goal (e.g., grades).
- Social conditions that affect motivation and well-being: these can affect "feelings of competence, autonomy, and connectedness," which aid learning.
- Setting and specifying goals: these aid learning, either by the instructor setting learning goals for students or students doing so for themselves.
- We value what we own or create: students are likely to value projects that they create.
- Social learning: belonging can increase student motivation; this is particularly important for stigmatized groups.

Another very useful source is [Ambrose et al. \(2010\)](#). While a book, its major points can be summarized as follows; each of the following points has a chapter devoted to it:<sup>18</sup>

- "Students' prior knowledge can help or hinder learning."
- "How students organize knowledge influences how they learn and apply what they know."
- "Students' motivation determines, directs, and sustains what they do to learn."
- "To develop mastery, students must acquire component skills, practice integrating them, and know when to apply what they have learned."
- "Goal-directed practice coupled with targeted feedback enhances the quality of students' learning."
- "Students' current level of development interacts with the social, emotional, and intellectual climate of the course to impact learning."
- "To become self-directed learners, students must learn to monitor and adjust their approaches to learning."

A final guide to findings from cognitive science used here is [Willingham \(2009\)](#). While he writes for the K-12 teachers (thus, "children" below), his points apply to higher education as well. His chapter summaries are:<sup>19</sup>

- "People are naturally curious, but not good thinkers; unless the cognitive conditions are right, we avoid thinking."
- "Factual knowledge must precede skill."
- "Memory is the residue of thought."
- "We understand most things in the context of things we already know, and most of what we know is concrete."
- "It is virtually impossible to become proficient at a task without extended practice."
- "Cognition early in training is fundamentally different from cognition late in training."
- "Children are more alike than different in how they think and learn."
- "Children do differ in intelligence, but it can be changed through sustained hard work."
- "Teaching, like any complex cognitive skill, must be practiced to be improved."

Some of these points might seem a bit cryptic and taken as a whole they may verge on daunting. Yet, they provide a very rich framework for thinking about teaching. The implementation of many of these points is

<sup>18</sup>The wording is from [Ambrose et al. \(2010, pp. 4-6\)](#).

<sup>19</sup>The wording comes from the book.



described next.

## 5 Redesigned Class Structure

### 5.1 Order of Topics

To investigate the impact of the concepts described above, a principles of macroeconomics course was redesigned by the second author and implemented at Penn State. The details below are for two sections during the fall of 2015. The combined final enrollment was 684 students.

In line with suggestions from [Bain \(2004\)](#), who found that the best college instructors oriented their courses around big questions, this course was structured around two major questions:

1. Why are some countries poor? / Why do some countries grow while others don't? / Will American real incomes continue to double each generation?
2. Why were so many unemployed, and why are so many currently underemployed and out of the labor force? What might be done, if anything?

This is also consistent with the point by [Willingham \(2009, Ch. 1\)](#) that people are naturally curious and this is used to help motivate learning.<sup>20</sup> Besides big questions for the course, every class or two a new puzzle is presented to help maintain curiosity.<sup>21</sup> These questions hopefully steer the class towards more of an “achievement motivation” ([Moulton, 2014](#)) and make learning more intrinsically motivated.

To answer these questions, standard tools common to macro principles are employed. The course is organized into five sections:

1. Measuring the Economy: GDP, Prices, and Inflation
2. Long Run Growth
3. Unemployment and the Business Cycle
4. Explaining the Economy's Movements in the Short Run (i.e., aggregate supply and demand)
5. Monetary and Fiscal Policy

While this is not a traditional sequence, the usual topics are covered and more importantly it leads to the quickest possible answers to the big course questions. Tools to answer these questions are developed just before they are needed. Question 1 is answered in Section 2, the first part of Question 2 is answered in Section 4, and the second part of Question 2 is resolved in Section 5.

In addition, in Section 1 there is some emphasis put on the behavior of key macro aggregates since 1970:<sup>22</sup> real GDP, the GDP deflator, and in Section 4 on the behavior of key labor market variables: the unemployment rate, employment, and the labor force participation rate, as well as an introduction to the business cycle. This is consistent with the point by [Willingham \(2009, Ch. 2\)](#) that factual knowledge must precede skills and that we understand concrete ideas (i.e., behavior of key macro variables) before abstract ones (aggregate supply and demand).

### 5.2 Student Preparation for Class: JiTTs

To make class time more productive, students are assigned textbook readings before class. This was enforced with “JiTT” assignments. Every two or three weeks as the class approached a new topic, students were assigned a “Just in Time Teaching” (JiTT) assignment in the course management system. Each JiTT is a short set of essay questions on assigned readings from the textbook. In this sense, the class is “flipped,” but this is not the defining characteristic of the course. JiTT questions are selected by the instructor to focus on topics that the instructor has found to be difficult for his or her students. As the topic has yet to be covered in class, they are lightly graded—students generally receive full credit as long as it is reasonably clear that they read the assignment

<sup>20</sup>This was first explored by the second author in [Goffe \(2007\)](#).

<sup>21</sup>Examples include: “How could GDP have risen during the 2001 recession?” (nominal, not real GDP grew) “In 1913, coffee cost \$.05, a bag of fries cost \$.10, and a silk scarf cost \$2. Have we been harmed by the increase in prices since then?” “Is economic growth across countries zero-sum?” “What average annual growth rate of the economy will make your life so radically different from the lives of your grandparents? 1%? 2%? 5%? 10%?”

<sup>22</sup>This date is a bit arbitrary but it does include a wealth of macroeconomic behavior.

(even if some details might be incorrect). To illustrate, here are the JiTT questions for the first part of the first section:

1. As you read, GDP includes some items and it excludes others. Do you agree or disagree with the choices that were made? Why or why not?
2. How are real and nominal GDP different? How are they similar?
3. What did you find confusing, interesting, or surprising in this reading?

The last question is used with every JiTT to help students consider what they understand or do not; this addresses the point made by [Ambrose et al. \(2010, Ch. 7\)](#) that students should monitor their own thinking—novices rarely do this, while experts do. The term for this is “metacognition” (or thinking about one’s own thinking) and this question is suggested in [Mazur and Watkins \(2009\)](#). This question also helps students evaluate if what they read was different than their prior thoughts and alerts instructors to this as well. This addresses the point raised by [Ambrose et al. \(2010, Ch. 1\)](#) on students’ prior thinking. Here are several responses to the last question:

1. “I found the circular flow diagram confusing. I also didn’t understand why the value of total production is equal to the value of total income.”
2. “I found the topics of real and nominal GDP to be pretty confusing.”
3. “Now when I see GDP pop up on CNBC, I understand how it is calculated, what it measures, and where those measurements come from.”

Instructors can find these responses particularly valuable as they identify problem areas for the class. That is, they can help address the curse of knowledge ([Moulton, 2014](#)).

Table 1 shows the percent of students who completed each of the five JiTT during the semester as fraction of those who finished the course. Keep in mind that the lowest JiTT score was dropped, so some missed JiTTs may simply have been strategic. These data suggest that it is possible to have students read the text.

Table 1: Percent of Each JiTT Completed by Section

| JiTT      | 1     | 2     | 3     | 4     | 5     |
|-----------|-------|-------|-------|-------|-------|
| Section 2 | 92.8% | 93.8% | 93.4% | 88.1% | 86.6% |
| Section 3 | 85.2% | 91.4% | 90.4% | 87.7% | 81.5% |

For more on JiTTs, see [Guertin et al. \(2012\)](#), [Simkins and Maier \(2009\)](#), and [Simkins \(2012\)](#).

### 5.3 Clicker Questions

While there was some lecture, a significant amount of class time was spent on clicker questions.<sup>23</sup> There were an even 200 questions over the course of the semester<sup>24</sup> making for roughly 14 questions per week, or a bit less than 5 per day as the sections met on Monday, Wednesday, and Friday. [Wieman et al. \(2017\)](#) was one guide for the development of the questions and the other was noticing what topics students had difficulty with (such as on exams or during off hours) and then a clicker question was used to probe that difficulty and ultimately help students deeply understand a topic.

The first six types of questions correspond to the more effective types of clicker questions described in [Wieman et al. \(2017\)](#). The next type, “Introduce a Concept,” is not explicitly mentioned but is in the same spirit as they typically involve deep processing by the student. Examples of each of these can be found in the Appendix. The final set of four types are less effective ([Wieman et al., 2017](#)), but they played a useful role. Non-economic survey questions involved setting the stage for the class the first day, like asking: “Do you know the name of the people sitting next to you?” (typically “no”) or asking if sound was adequate. Instead of reading the syllabus, questions were asked about it the second day of class. There were some calculations and finally students were surveyed as to

<sup>23</sup>The campus standard iClicker was used and it was found to be very user-friendly for students, and the instructor as well. It was also very stable and reliable. A key iClicker advantage is that the instructor can see vote totals for the class independent of the class. This avoid tipping off students about their collective voting.

<sup>24</sup>This is actually for Section 2 as some clicker data for Section 3 was lost, but the two sections were run in a nearly identical fashion.

Table 2: Types of Clicker Questions

|                                |    |
|--------------------------------|----|
| Connections to Prior Knowledge | 4  |
| Different Representations      | 6  |
| Survey                         | 10 |
| ConcepTest                     | 14 |
| Interpret a Chart of Data      | 15 |
| Apply a Concept                | 93 |
| Introduce a Concept            | 39 |
| Non-Econ Survey                | 4  |
| Syllabus                       | 8  |
| Calculation                    | 12 |
| Quiz                           | 13 |

which quiz questions they would like a detailed explanation for.<sup>25</sup> “Peer Instruction,” popularized by Eric Mazur (Crouch et al., 2007) and studied in more detail by Smith et al. (2011) was used for 51 of the questions.<sup>26</sup> That is, a question was posed and if the class was split, with typically between 30% and 70% of the class answering correctly (students did not see the results of the class vote as iClicker provides a receiver that shows the vote totals to the instructor) students were then asked to “convince their neighbor of their answer.” This usually took 2–3 minutes and the instructor talked to various students to understand their thinking—this helps address the “curse of knowledge” (Moulton, 2014) and to also understand their prior knowledge (Ambrose et al., 2010, Ch. 1). Almost always, the class converged to the correct answer based solely on student interaction. Figure 7 shows the distribution of improvement in the votes in Section 2. Note that the mean improvement in students answering correctly was 21.6 (if 50% of the class was correct on the first vote, 71.6% were correct on the second one). After the second vote and following Smith et al. (2011), the instructor thoroughly described the correct answer and the reasoning behind it.

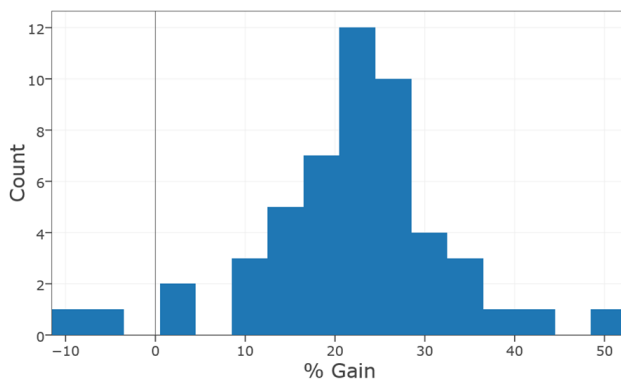


Figure 7: Distribution of changes in correct votes during Peer Instruction in Section 2.

Of the non-survey questions, 54.4% of questions were answered correctly. This illustrates “desirable difficulties” (Moulton, 2014)—ideas “stick better” if the work is challenging. These clicker questions also invoke deliberate practice (Deslauriers et al., 2011; Moulton, 2014), retrieval practice (Moulton, 2014), social learning (Moulton, 2014) with Peer Instruction, “memory is the residue of thought” (Willingham, 2009, Ch. 3), strengthening of

<sup>25</sup>A small number of course points were awarded for clicker participation but no credit was given for correct answers. James (2006) finds that high stakes grading of clicker questions leads to lower quality student discussion during Peer Instruction.

<sup>26</sup>As Peer Instruction involves two clicker questions, half of class clicker questions were part of Peer Instruction since 200 questions were asked over the semester.

students' schemata of course topics (Ambrose et al., 2010, Ch. 2) with questions that asked about relationships between topics, and addressing prior knowledge (Ambrose et al., 2010, Ch. 1). Also, many of the questions offer concentrated practice on challenging concepts—like how a coach will work on part of a golf swing as opposed to watching a player in a tournament.<sup>27</sup>

## 5.4 Worksheets

Since many of the clicker questions are conceptual in nature, students receive little in-class feedback on performing actual calculations or in drawing graphs. To address this issue five worksheets were used over the semester. Topics included calculating a price index, calculating rates of growth of prices and real GDP, manipulating a per-worker production function, and shifting aggregate supply and demand curves and finding the resulting equilibria. Students spent several minutes on the worksheets and then conferred with peers to answer clicker questions on worksheet results. The instructor and the learning assistants checked with students as they worked through the questions to understand their thinking. After the clicker questions were done, the instructor worked through the answers to all questions on a document camera. See Jones et al. (2015) for a thorough description of how worksheets and the principle of deliberate practice improved learning in a senior-level physics course.

## 5.5 Lecture

While considerable time in class was devoted to clicker questions on topics that students read outside of class, there was still an appreciable amount of lecture.<sup>28</sup> This is particularly true for topics not covered in the text. One particular relevant finding from cognitive science to lecture are the constraints of human working memory. Moulton (2014) puts it this way: “Our minds have severe, inherent limits on the amount of information they can consciously process at any given time and this limit varies dramatically based on our experience with and strategies for processing that information.” Meanwhile, Deans for Impact (2015) use this phrasing: “Students have limited working memory capacities that can be overwhelmed by tasks that are cognitively too demanding.” To deal with these limitations during lecture, an outline was provided for each class meeting and it was projected on an auxiliary screen. It also included key abbreviations and definitions. PowerPoint slides were structured so that instead of all slide content appearing at once, words came in line by line or, in the case of graphs, curve by curve. Clicker questions, which were on the slides, were not read to the students, but left for the students to read to themselves. During class, pains were made to connect new ideas to previous ones and thus connect to long-term memory. By making connections, it was hoped that students would thus have a more complex “schema” (or structure of knowledge) of course topics. In addition, clicker questions hopefully helped students consolidate ideas in working memory into long-term memory. Resources along these lines include Wieman (2015) and Fenesi (2015); in part, both summarize the work of the cognitive scientist Richard E. Mayer for the classroom setting. Finally, where possible, concrete examples preceded abstract ideas (Willingham, 2009, Ch. 4).

## 5.6 Quizzes and Exams

Rather than the traditional two or perhaps three midterms, evaluations during the semester was with a series of seven quizzes. The questions were exam-caliber but there were fewer questions. The total number of questions on the quizzes was roughly comparable to the total number of exam questions if there were two midterms. Effectively, then, what would have been midterms were spaced out over time. This led to following advantages:

1. As each quiz only took a portion of the class time, there was time to go over the quizzes immediately after they were completed. The answers were revealed and after consulting with their peers, students voted on the questions they wished to see explained in more detail. The combination of feedback and thinking about what they do not understand, after consultation with their peers, constitute deliberate practice and metacognition.

---

<sup>27</sup>An alternative to JiTTs and Peer Instruction is “Team-Based Learning.” It is widely used in the health sciences, where learning is paramount.

<sup>28</sup>Note that lecture, technically known as “direct instruction,” can be effective in some situations; see Schwartz and Bransford (1998). However, this does not include an instructor talking the length of a class.

2. After the first quiz, each succeeding one contained questions on topics from previous quizzes (all topics on quizzes were announced). This is an example of “spaced practice” that [Moulton \(2014\)](#) made—learning is retained better if the topic is repeated over time.

Another way to think of this assessment structure is that it combines formative assessment with summative assessment.

Some quizzes were multiple choice and others were short-answer. It should be emphasized that most of the multiple choice questions were fairly high up on Bloom’s Taxonomy—typically at the “apply” or “analyze” level. When students were asked about what advice they would give students in later semesters, a fairly typical response was “I would recommend to take good notes and be able to apply the concepts that are covered in class to be able to answer the questions. The questions on the quizzes are usually not just straight memorization and you have to be able to apply the concepts.”

A clear cost to the instructor of this quizzing approach is that since each section left class with a copy of the quiz, new ones must be written for each semester and each section. Additionally, if a question is of poor quality students will likely point it out then and there, which can be unsettling to the instructor.

The spacing effect was again utilized with a comprehensive final exam.

## 5.7 Quiz Reflections

About a week after each quiz there was a brief “quiz reflection”<sup>29</sup> carried out in the course management system. To aid student metacognition ([Ambrose et al., 2010](#), Appendix F), students were asked brief questions of two types. First, they were asked which quiz question they found most difficult on the previous quiz, if they now understood it, and what they would do if they did not fully understand it. The topics students found most difficult frequently surprised the instructor; this would seem to be an example of the curse of knowledge ([Moulton, 2014](#)) and those topics were typically addressed in a later homework. Second, students were asked about how they studied and how they might improve their study methods. [Dunlosky et al. \(2013\)](#) was used for study suggestions; note that [Dunlosky \(2013\)](#) summarizes these research-based findings in a format suitable for students. Another option for students is [Chew \(2015\)](#). It is worth noting that the author of these videos, Stephen Chew, is a cognitive scientist and in 2011 was a “U.S. Professor of the Year,” perhaps the highest award given in the U.S. for college teaching. His award is partly based on these videos. A more succinct option is [Sumeracki et al. \(2017\)](#).

## 5.8 Homeworks

There were five homeworks over the course the semester; all questions were written by the instructor. Topics were shuffled to achieve “interleaving” ([Moulton, 2014](#)) and each homework after the first one contained a significant number of questions from previous parts of the course to implement spacing. [Butler et al. \(2014\)](#) investigated a modified homework structure by adding spacing, retrieval practice, and feedback (students had to view the homework key to receive credit for the homework). They found that students scored about a letter grade higher on exam questions on topics that used this intervention. In addition, [Wieman \(2016\)](#)<sup>30</sup> was followed as much as the course allowed: for calculations students had to select the appropriate data (they were given more than needed), were asked to make qualitative judgements on an answer being reasonable, and identify what concepts needed to be used. These clearly involve deep processing ([Moulton, 2014](#)) as well as acquiring “component skills, practice integrating them, and know when to apply what they have learned” ([Ambrose et al., 2010](#), Ch. 4). Hopefully the questions also enhanced the generation of more complex schemata ([Ambrose et al., 2010](#), Ch. 2) as some questions explicitly asked about relationships between topics.

---

<sup>29</sup>These are sometimes called “exam wrappers” as they are completed with or soon after an exam and passed back to students before the next exam.

<sup>30</sup>This guidance can be seen as an elaboration of [Bangs et al. \(2012\)](#).

## 6 Empirical Results

Two sections of the second author's principles of macroeconomics were taught with the above methods in the fall of 2015. 636 students took the final exam and the TUCE was offered as a bonus on it. Students could earn 1,000 points in the course<sup>31</sup> and 1/3 of a point was awarded for each TUCE question answered correctly. 508 of these students also took the TUCE at the start of the semester for a bonus of 10 points for completing the TUCE; it seemed inappropriate to offer a bonus for correct answers. The campus testing center was utilized to avoid using class time.

Thus, there is pre-post TUCE data for 80% of those who completed the course. The post-TUCE results are basically the same for those who took the pre-TUCE and those who did not. The mean pre-TUCE score was 11.90 questions correct (out of 30) and the mean post-TUCE score was 19.91. The gain,  $g$ , was .44. Recall that in the TUCE norming population, the mean pre-TUCE score was 9.80 questions correct and the mean post-TUCE score was 14.19, for a gain of .22. Thus, the students in this study had twice the gain of the norming population.

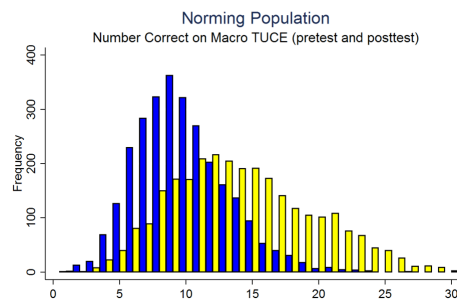


Figure 8: Blue denotes the number of correct answers on the pretest and yellow on the post-TUCE for students in the TUCE norming population.

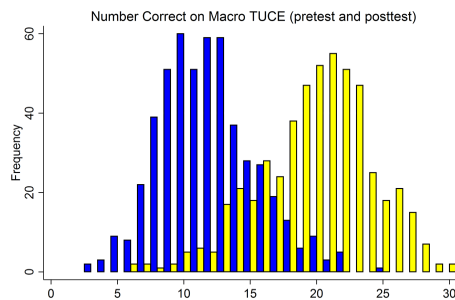


Figure 9: Blue denotes the number correct of correct answers on the pretest and yellow on the post-TUCE for students in the redesigned course described in this paper.

Figure 8 shows the pre- and post-TUCE norming population distributions (blue and yellow, respectively) while Figure 9 shows the same distributions for the students in this study. Improvement on the post-TUCE for students in the redesigned class was 3.71 points higher (on average) than students in the norming population. This

<sup>31</sup>592 points could be earned with quizzes, 200 points for the final, 80 points on the JiTTs, 80 points for homeworks, 40 points for clicker participation, and 8 for the quiz reflections. The lowest quiz, JiTT, homework, and quiz reflection were automatically dropped.

represents .77 standard deviations of the norming population improvement. The 99.5% confidence interval is from 3.16 to 4.26.

One notable difference between the norming population and the students in the redesigned class in this study is the pre-TUCE mean scores: 9.80 versus 11.90. Perhaps students with a higher pre-TUCE score are capable of learning more. Thus, data from the norming population was used to generate a synthetic set of data (call it the "norming *sample*") matching the pre-TUCE distribution seen in the redesigned course described in this paper. Figure 10 shows the distributions: blue shows the pre-TUCE distribution for both the redesigned course and the norming sample, red shows the post-TUCE distribution of the norming sample while green is the post distribution of students in the redesigned course. The difference in the means is 3.93 with a difference of .82 standard deviations. The 99.5% confidence interval for this difference is from 3.20 to 4.66.

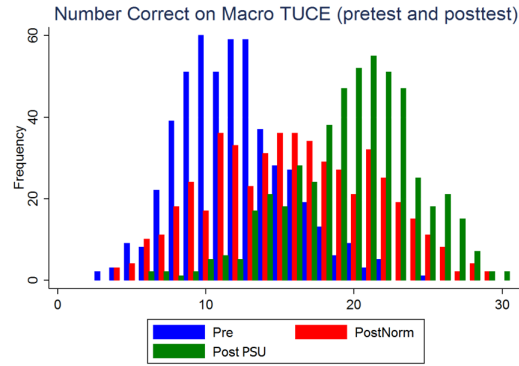


Figure 10: The synthetic sample is in blue (pre-TUCE) and red (post-TUCE). Green denotes post-TUCE scores from the redesigned course.

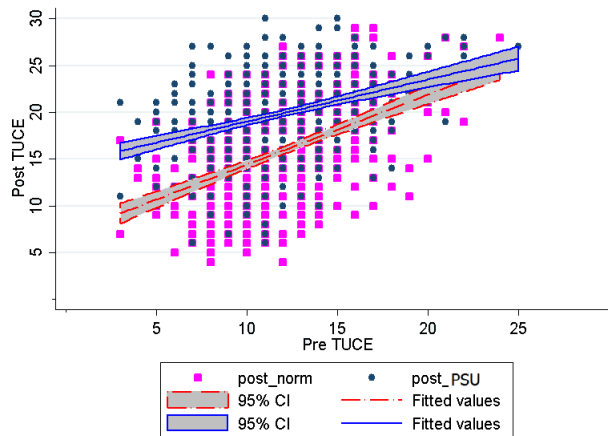


Figure 11: Pre- and post-TUCE scatter plot with a synthetic distribution (pink) and for students in the redesigned course (blue).

Finally, Figure 11 compares the matched distribution from the norming sample (pink) and students in re-

designed course (blue). Note how for a given pre-TUCE score, students in the redesigned course generally had higher post-TUCE scores. The difference is statistically significant for the 97% of students who scored below 20 on the pre-TUCE.

## 7 Conclusion

The empirical results in this study suggest that using a research-based (or “evidence-based”) approach to teaching principles of macroeconomics might be a fruitful endeavor. The results presented in the last section were both statistically significant and showed a meaningful learning increase compared to what is typically seen on the Test of Understanding of College Economics (TUCE) with an effect size of approximately .8 standard deviations. While the gains in this intervention fell short of the roughly two standard deviations seen in [Hake \(1998\)](#) and [Deslauriers et al. \(2011\)](#), recall [Hake \(1998\)](#) used an instrument well-tuned to common student misconceptions that lecture generally does not dispel. [Deslauriers et al. \(2011\)](#) only looked at the classroom component of learning and was not influenced by students’ studying.

Besides rather promising empirical gains, research-based teaching methods provide instructors with a robust framework to view teaching and learning. Indeed, findings from [Andrews et al. \(2011\)](#) suggest that a “rich and nuanced understanding of teaching and learning” and not active learning per se are an essential component of improving student learning.

An additional benefit to these methods is the instructor is almost certain to better understand students’ prior knowledge, which can be a key impediment to learning. Writing JiTT and clicker questions combined with reviewing the answers almost automatically forces the instructor to reflect upon difficult topics for students and how they can be presented. Thus, the curse of knowledge is addressed multiple times per week of instruction.

In this study numerous teaching methods that seem to be fairly uncommon in economics classrooms were implemented. Given the experimental design, it is difficult to say which ones might be most important; indeed, it is plausible that they acted synergistically. That said, four innovations seem to stand out. First, students received considerable feedback during the semester. They were asked two hundred challenging clicker questions of the types that STEM researchers have found to be particularly effective and they were asked and answered in ways that have been found to be most useful; these include Peer Instruction. In addition, in lieu of midterms, students took seven quizzes; effectively, midterms were spread out over time. After the quizzes were turned in, the answers and the reasoning behind them were discussed. These approaches to clicker and quiz questions implement three principles that cognitive scientists have found that aid learning: deliberate practice (asking challenging questions and following up promptly with complete answers), retrieval practice (recalling facts and principles which aids their retention) and desirable difficulties (we are more likely to remember what took some effort). Second, there was considerable spacing of learning—students were tested on many topics several times over the semester. Third, many of the clicker questions and some of the quiz questions were conceptual rather than computational so they were higher on Bloom’s Taxonomy. That is, instead of doing problems, students were often asked about fundamental concepts that underlie problem solving. A key finding from STEM education research is that students can often solve problems with scant fundamental understanding of key concepts. Such questions should help students develop richer schemas (networks of knowledge). Fourth, quiz wrappers (taken after quizzes) and JiTT questions include questions where students were asked to assess their understanding of a topic, or their metacognition. This should aid the development of students’ schemas.

Perhaps in the future economic education researchers might undertake more studies employing ideas and concepts from STEM education research and cognitive science to help improve the modest learning in economics classrooms outlined above in Section 2. There are certainly many opportunities for further research. Our students would be the beneficiaries.



## Appendix

Section 5.3 briefly described the different types of clicker questions used in this study and this appendix contains some actual examples. They are considered by [Wieman et al. \(2017\)](#) to be more effective types of clicker questions but for “introducing a topic,” and that one is similar in spirit to the others as it requires deep thinking. All student responses are from Section 2 of the study as some clicker responses were lost from Section 3. Figure 12 demonstrates a question that asks students to compute real GDP. Note that there was no lecture on this topic—they were asked this question after completing the first JiTT, which included readings on real and nominal GDP. The caption includes both votes—the first is students individually answering and the second is after students try to convince each other of the correct answer (i.e., “Peer Instruction”) and then vote again. Note how the percent of the class answering correctly went from 62% to 85%. In effect, students were teaching each other in a carefully controlled situation. Students who understood the topic had their understanding reinforced and students who did not had it explained to them by someone who just learned it, who might be the best teacher as they understood the pitfalls better than an instructor who likely has known this material for years. Following [Smith et al. \(2011\)](#), after all voting, the thinking behind the answer was described to the class by the instructor. This includes students’ thoughts on the question gather by the instructor asking students for their understanding when students are conferring with each other during Peer Instruction.

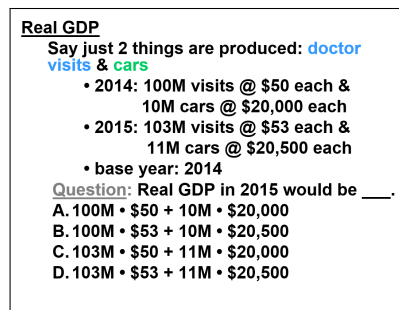


Figure 12: A: 9%, B: 7%, C: 62%, D: 21% to  
 A: 3%, B: 4%, C: 85%, D: 9%

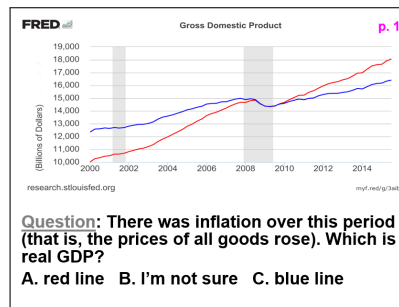


Figure 13: A: 34% B: 12%, C: 54% to A: 16%, B: 3%, C: 80%

The question in Figure 13 followed the previous one. It illustrates how students might be able to answer a question with one representation of an idea, like calculating real and nominal GDP, but then have difficulty with a “different representation” of that concept—here with a time series plot of real and nominal GDP. It also asks students to interpret a chart of data, another effective type of clicker question. On the first vote, only a bare majority, 54%, knew the correct answer, but after conferring with their peers, 80% did. The “I’m not sure

response” was designed to have students think metacognitively about this question—did they understand it or not? This also illustrates deepening students’ understanding of real and nominal GDP; i.e., enriching their schema (Ambrose et al., 2010, Ch. 2).

The question in Figure 14 investigates students’ prior knowledge (Ambrose et al., 2010, Ch. 1). Note how student thinking is considerably different from how economists think of this topic. As STEM researchers have found that just telling students that they are wrong can lead to little learning, an “Interactive Lecture Demonstration” (Simkins and Maier, 2008) was used. After eliciting student thinking with this question, students watched Rosling (2010), a 4-minute summary of changes in real per capita GDP and life expectancy for the last two centuries by country using the well-know data package Gapminder. After the video they were then asked to vote again and a discussion ensued. Given the questions that are typically raised, students seem to have difficulty with the concept that growth is not zero-sum.<sup>32</sup>

**Economic Growth: Zero-Sum?**  
**Question:** In your view, are some countries rich because they have taken wealth, income, and resources from poor countries (i.e., is world growth zero-sum)?  
A. yes  
B. mostly yes  
C. it is an even mix of factors  
D. mostly no  
E. no

Figure 14: A: 21%, B: 22%, C: 36%, D: 12%, E: 10%

The question in Figure 15 is an example of a survey question. It is used the first day of class and the results are often referred to any time unemployment is mentioned to connect pre-existing student thinking with course topics (Ambrose et al., 2010, Ch. 1).

**Survey:** How many people do you know who are or in the last few years were  
- unemployed (not working but looking)  
or  
- underemployed (working part time but wish to work full time)  
or  
- have given up looking for a job:  
A. 0 B. 1 C. 2 D. 3 E. 4 or more

Figure 15: A: 19%, B: 19%, C: 26%, D: 12%, E: 25%

The question in Figure 16 is an example of a conceptual question; these are known in STEM disciplines as a “conceptest.” Note how it does not ask for a calculation but instead requires a reasonable understanding of the topic; in this case, the difference between the price level and inflation, which is often a confusing point for principles students, as the first vote attests. Note the large movement to the correct answer with the second vote. The instructor explanation after all voting included a numerical example. Note that option C asks students to judge their understanding—i.e., use metacognition.

<sup>32</sup>Of course, it would be helpful if economists had a list of common misconceptions, as physicists do. Bice et al. (2014) is an attempt in that direction.

**Inflation vs. the Price Level**  
**Question:** Say that inflation fell. Then the price level (measured by the CPI or GDP deflator) would be sure to fall as well.  
 A. true (I'm sure)  
 B. true (I think)  
 C. not sure  
 D. false (I think)  
 E. false (I'm sure)

Figure 16: A: 37%, B: 19%, C: 4%, D: 9%, E: 30% to  
 A: 24%, B: 7%, C: 2%, D: 8%, E: 59%

The question in Figure 17 is an example of students being asked to apply a concept. Again, after conferring with their peers, the class moves to the correct answer.

**Real vs. Nominal GDP**  
**Question:** In a recession, production declines. Also, the U.S. almost always has inflation. With these in mind, which is most likely to fall in a recession?  
 A. real GDP  
 B. nominal GDP  
 C. I'm not sure

Figure 17: A: 52%, B: 46%, C: 2%, to A: 74%, B: 25%, C: 0%

Finally, the question in Figure 18 is an example of introducing a concept.<sup>33</sup> Instead of first giving students a formula to calculate the change in a real price, this question introduces this concept conceptually. The motivation behind this approach is what physics education researchers found in [Hestenes et al. \(1992\)](#)—students can correctly solve problems without a fundamental understanding of the relevant concepts.

**Real Prices: Subtraction**

| <u>date</u> | <u>CPI</u> | <u>nominal price of gasoline</u> |
|-------------|------------|----------------------------------|
| 3/2006      | 200        | \$2.41                           |
| 1/2011      | 220        | \$3.06                           |

**Question:** Over this period, the real price of gas \_\_\_ (i.e. price after taking out inflation). Hint: which grew faster?  
 A. rose B. stayed the same  
 C. fell D. I'm not sure

Figure 18: A: 58%, B: 29%, C: 7%, D: 6%

<sup>33</sup>Those are actual values for the CPI and the national nominal average price of gasoline for those dates; the dates were chosen so that the inflation rate can be easily computed.

## References

- Allgood, Sam, William B. Walstad, and John J. Siegfried**, "Research on Teaching Economics to Undergraduates," *Journal of Economic Literature*, 2015, 53 (2), 285–325. <<https://www.aeaweb.org/articles?id=10.1257/jel.53.2.285>>. 5
- Ambrose, Susan A., Michael W. Bridges, and Michele DiPietro**, *How Learning Works: Seven Research-Based Principles for Smart Teaching*, Jossey-Bass, 2010. <<http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470484101.html>>. 1, 5, 10, 12, 13, 14, 15, 20
- Andrews, T.M., M.J. Leonard, C.A. Colgrove, and S.T. Kalinowski**, "Active Learning Not Associated With Student Learning in a Random Sample of College Biology Courses," *CBE-Life Sciences Education*, 2011, 10 (4), 394–405. <<http://www.lifescied.org/content/10/4/394.long>>. 18
- Bain, Ken**, *What the Best College Teachers Do*, Harvard University Press, 2004. 11
- Bangs, Joann, Jennifer Docktor, Ken Heller, Brian Peterson, and Rochelle Ruffer**, "Context-Rich Problems," 2012. <[http://serc.carleton.edu/econ/context\\_rich/index.html](http://serc.carleton.edu/econ/context_rich/index.html)>. 15
- Bice, Douglas, Elisabeth S. Curtis, Wayne Geerling, William L. Goffe, Adam Hoffer, Solina Lindahl, Mark H. Maier, Brian Peterson, and Wendy Stock**, "Preconceptions of Principles Students," 2014. <<http://cook.rfe.org>>. 2, 20
- Bransford, J.D., A.L. Brown, R.R. Cocking et al.**, *How People Learn*, National Academy Press Washington, DC, 2000. <[https://download.nap.edu/login.php?record\\_id=9853&page=%2Fcatalog.php%3Frecord\\_id%3D9853](https://download.nap.edu/login.php?record_id=9853&page=%2Fcatalog.php%3Frecord_id%3D9853)>. 1, 5
- Brown, Peter C, Henry L. Roediger III, and Mark A. McDaniel**, *Make It Stick: The Science of Successful Learning*, Harvard University Press, 2014. <<http://www.hup.harvard.edu/catalog.php?isbn=9780674729018>>. 1, 5, 9
- Butler, Andrew C., Elizabeth J. Marsh, JP Slavinsky, and Richard G. Baraniuk**, "Integrating Cognitive Science and Technology Improves Learning in a STEM Classroom," *Educational Psychology Review*, 2014, 26 (2), 331–340. <<http://link.springer.com/article/10.1007%2Fs10648-014-9256-4>>. 15
- Chew, Stephen**, "How to Get the Most Out of Studying Video Series," 2015. <<https://www.youtube.com/playlist?list=PL85708E6EA236E3DB>>. 15
- Concept Inventories/Conceptual Assessments in Biology (CABs)**, 2017. <<http://go.sdsu.edu/dus/ctl/cabs.aspx>>. 4
- Crouch, Catherine H., Jessica Watkins, Adam P. Fagen, and Eric Mazur**, "Reviews in PER Volume 1: Research-Based Reform of University Physics," 2007. <<http://www.compadre.org/Repository/document/ServeFile.cfm?DocID=241&ID=4990>>. 6, 13
- Deans for Impact**, *The Science of Learning* 2015. <[http://www.deansforimpact.org/the\\_science\\_of\\_learning.html](http://www.deansforimpact.org/the_science_of_learning.html)>. 5, 14
- Deslauriers, Louis, Ellen Schelew, and Carl Wieman**, "Improved Learning in a Large-Enrollment Physics Class," *Science*, 2011, 332 (6031), 862–4. <[http://www.cwsei.ubc.ca/SEI\\_research/index.html](http://www.cwsei.ubc.ca/SEI_research/index.html)>. 1, 2, 6, 7, 13, 18
- Dunlosky, John**, "Strengthening the Student Toolbox: Study Strategies to Boost Learning.," *American Educator*, 2013, 37 (3), 12–21. <<http://www.aft.org/periodical/american-educator/fall-2013/strengthening-student-toolbox>>. 15

- , **Katherine A Rawson, Elizabeth J Marsh, Mitchell J Nathan, and Daniel T Willingham**, “Improving Students’ Learning with Effective Learning Techniques Promising Directions from Cognitive and Educational Psychology,” *Psychological Science in the Public Interest*, 2013, 14 (1), 4–58. <<http://psi.sagepub.com/content/14/1/4.abstract>>. 15
- Ericsson, K. Anders and Robert Pool**, *Peak: Secrets from the New Science of Expertise*, Houghton Mifflin Harcourt, 2016. <<http://www.hmhc.com/shop/books/Peak/9780544456235>>. 1, 6
- , **Ralf Th. Krampe, and Clemens Tesch-Romer**, “The Role of Deliberate Practice in the Acquisition of Expert Performance,” *Psychological Review*, 1993, 100 (3), 363–406. <<http://projects.ict.usc.edu/itw/gel/EricssonDeliberatePracticePR93.pdf>>. 6
- Fenesi, Babara**, “Effective Multimedia Design,” January 2015. <<https://www.youtube.com/watch?v=0I3v7of9dEc>>. 14
- Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth**, “Active Learning Increases Student Performance in Science, Engineering, and Mathematics,” *Proceedings of the National Academy of Sciences*, 2014, 111 (23), 8410–8415. <<http://www.pnas.org/content/111/23/8410.abstract>>. 5
- Goffe, William L.**, “Redesigning a Large Macro Principles Course Using Bains What the Best College Teachers Do,” 2007. <<http://cook.rfe.org>>. 11
- , “Initial Misconceptions in Macro Principles Classes,” 2013. <<http://cook.rfe.org>>. 2
- Guertin, Laura, Carol Ormand, Gregor Novak, Andy Gavrin, Scott P. Simkins, Gregor Novak, Marcelo Clerici-Arias, and Rae Jean Goodman**, “Just in Time Teaching (JiT),” 2012. <<http://serc.carleton.edu/econ/justintime/>>. 12
- Hake, Richard R.**, “Interactive-Engagement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses,” *American Journal of Physics*, 1998, 66, 64–74. <[https://www.researchgate.net/profile/Richard\\_Hake/publication/228710512\\_Interactive-engagement\\_versus\\_traditional\\_methods\\_A\\_six-thousand-student\\_survey\\_of\\_mechanics\\_test\\_data\\_for\\_introductory\\_physics\\_courses/links/0912f50fa0de905573000000.pdf](https://www.researchgate.net/profile/Richard_Hake/publication/228710512_Interactive-engagement_versus_traditional_methods_A_six-thousand-student_survey_of_mechanics_test_data_for_introductory_physics_courses/links/0912f50fa0de905573000000.pdf)>. 1, 2, 4, 5, 6, 9, 18
- Hartle, R. Todd, Sandhya Baviskar, and Rosemary Smith**, “A Field Guide to Constructivism in the College Science Classroom: Four Essential Criteria and a Guide to Their Usage,” *Bioscene: Journal of College Biology Teaching*, 2012, 38 (2), 31–35. <[http://www.acube.org/files/7513/6692/2474/Bioscene\\_December\\_2012\\_FINAL.pdf](http://www.acube.org/files/7513/6692/2474/Bioscene_December_2012_FINAL.pdf)>. 6
- Henderson, Charles, José P. Mestre, and Linda L. Slakey**, “Cognitive Science Research Can Improve Undergraduate STEM Instruction What Are the Barriers?,” *Policy Insights from the Behavioral and Brain Sciences*, 2015, 2 (1), 51–60. <<http://bbs.sagepub.com/content/early/2015/08/12/2372732215601115.full.pdf>>. 9
- , **Melissa Dancy, and Magdalena Niewiadomska-Bugaj**, “Use of Research-Based Instructional Strategies in Introductory Physics: Where Do Faculty Leave the Innovation-Decision Process?,” *Physical Review Special Topics-Physics Education Research*, 2012, 8 (2), 020104. <<http://link.aps.org/pdf/10.1103/PhysRevSTPER.8.020104>>. 5
- Hestenes, David, Malcolm Wells, and Gregg Swackhamer**, “Force Concept Inventory,” *The Physics Teacher*, 1992, 30 (3), 141–158. <<http://modeling.asu.edu/R&E/FCI.PDF>>. 2, 4, 5, 21
- Hoellwarth, Chance and Matthew J. Moelter**, “The Implications of a Robust Curriculum in Introductory Mechanics,” *American Journal of Physics*, 2011, 79 (5), 540–545. <[http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1332&context=phy\\_fac](http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1332&context=phy_fac)>. 5

- Hutson, Shane**, “Shane Hutson Interview,” in Trina McMahon and Derek Bruff, eds., *An Introduction to Evidence-Based Undergraduate STEM Teaching*, CIRT (Center for Integrated Research Teaching and Learning), 2014. <<https://www.youtube.com/watch?v=1KZ6HBY3tHo>>. 6
- James, Mark C.**, “The Effect of Grading Incentive on Student Discourse in Peer Instruction,” *American Journal of Physics*, 2006, 74 (8), 689–691. <<http://scitation.aip.org/content/aapt/journal/ajp/74/8/10.1119/1.2198887>>. 13
- Jones, David J., Kirk W. Madison, and Carl E. Wieman**, “Transforming a Fourth Year Modern Optics Course Using a Deliberate Practice Framework,” *Physical Review Special Topics-Physics Education Research*, 2015, 11 (2), 020108. <<http://link.aps.org/pdf/10.1103/PhysRevSTPER.11.020108>>. 14
- Mazur, Eric**, “Education: Farewell, Lecture?,” *Science*, 2009, 323 (5910), 50–51. <[mazor.harvard.edu/publications.php?function=display&rowid=635](http://mazor.harvard.edu/publications.php?function=display&rowid=635)>. 6
- **and Jessica Watkins**, “Just-in-Time Teaching and Peer Instruction,” in Scott P. Simkins and Mark H. Maier, eds., *Just in Time Teaching: Across the Disciplines, and Across the Academy*, Stylus Publishing, 2009. <<https://sty.presswarehouse.com/books/BookDetail.aspx?productID=178290>>. 12
- Miller, Michelle D.**, *Minds Online: Teaching Effectively with Technology*, Harvard University Press, 2014. <<http://www.hup.harvard.edu/catalog.php?isbn=9780674368248>>. 1, 5
- Moulton, Samuel T.**, *Applying Psychological Science to Higher Education: Key Findings and Open Questions* The Harvard Initiative for Learning and Teaching April 2014. <[http://hilt.harvard.edu/files/hilt/files/moulton\\_2014\\_applying\\_psychological\\_science\\_to\\_higher\\_education\\_april16.pdf](http://hilt.harvard.edu/files/hilt/files/moulton_2014_applying_psychological_science_to_higher_education_april16.pdf)>. 5, 9, 11, 12, 13, 14, 15
- PhysPort: Browse Assessments**, 2017. <<https://www.physport.org/assessments/>>. 4
- Resnick, Lauren B.**, “Mathematics and Science Learning: A New Conception,” *Science*, April 29 1983, 220 (4596), 477–78. <<http://www.sciencemag.org/content/220/4596/477.abstract>>. 5, 7
- Rosling, Hans**, “Hans Rosling’s 200 Countries, 200 Years, 4 Minutes - The Joy of Stats - BBC Four,” 2010. <<https://www.youtube.com/watch?v=jbkSRLYSojo>>. 20
- Schwartz, Daniel L. and John D. Bransford**, “A Time for Telling,” *Cognition and Instruction*, 1998, 16 (4), 475–5223. <[http://aaalab.stanford.edu/papers/A\\_time\\_for\\_telling.pdf](http://aaalab.stanford.edu/papers/A_time_for_telling.pdf)>. 14
- Simkins, Scott P.**, “Using Just-in-Time Teaching to Promote Student Learning in Economics,” in Gail M. Hoyt and KimMarie McGoldrick, eds., *International Handbook On Teaching And Learning Economics*, Edward Elgar, 2012. <<http://www.e-elgar.com/shop/international-handbook-on-teaching-and-learning-economics>>. 12
- **and Mark H. Maier**, “Learning from Physics Education Research: Lessons for Economics Education,” *Economics Research Network*, 2008. <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1151430](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1151430)>. 1, 20
- **and —**, *Just in Time Teaching: Across the Disciplines, and Across the Academy*, Stylus Publishing, 2009. <<https://sty.presswarehouse.com/books/BookDetail.aspx?productID=178290>>. 12
- Smith, M.K., W.B. Wood, K. Krauter, and J.K. Knight**, “Combining Peer Discussion with Instructor Explanation Increases Student Learning from In-Class Concept Questions,” *CBE-Life Sciences Education*, 2011, 10 (1), 55–63. <<http://www.lifescied.org/content/10/1/55.full>>. 8, 13, 19
- Sumeracki, Megan, Yana Weinstein, and Oliver Caviglioli**, “Six Strategies for Effective Learning,” December 2017. <<http://www.learningscientists.org/posters>>. 15

- Walstad, William B. and Jamie Wagner**, “The Disaggregation of Value-Added Test Scores to Assess Learning Outcomes in Economics Courses,” *The Journal of Economic Education*, 2016, 47 (2), 121–131. <<http://www.tandfonline.com/doi/abs/10.1080/00220485.2016.1146104>>. 1
- **and Sam Allgood**, “What Do College Seniors Know about Economics?,” *The American Economic Review*, May 1999, 89 (2), 350–354. <<http://www.jstor.org/stable/117134>>. 2, 4
- , **Michael W. Watts, and Ken Rebeck**, *Test of Understanding in College Economics: Examiner’s Manual*, fourth ed., National Council On Economic Education, 2007. 2
- Watts, Michael and Georg Schaur**, “Teaching and Assessment Methods in Undergraduate Economics: A Fourth National Quinquennial Survey,” *The Journal of Economic Education*, 2011, 42 (3), 294–309. <<http://www.tandfonline.com/doi/full/10.1080/00220485.2011.581956>>. 5
- Wieman, Carl**, “Why Not Try a Scientific Approach to Science Education?,” *Change*, September/October 2007, 33 (40), 9–15. <[http://www.cwsei.ubc.ca/resources/files/Wieman-Change\\_Sept-Oct\\_2007.pdf](http://www.cwsei.ubc.ca/resources/files/Wieman-Change_Sept-Oct_2007.pdf)>. 7
- , “Improving Learning by Reducing Unnecessary Mental Load,” January 2015. <[http://cwsei.ubc.ca/resources/files/Cognitive-Load\\_CWSEI.pdf](http://cwsei.ubc.ca/resources/files/Cognitive-Load_CWSEI.pdf)>. 14
- , “Creating Good Homework Problems (and Grading Them),” September 2016. <[http://www.cwsei.ubc.ca/resources/files/HomeworkProblems\\_CWSEI.pdf](http://www.cwsei.ubc.ca/resources/files/HomeworkProblems_CWSEI.pdf)>. 15
- , **Katherine Perkins, Sarah Gilbert, Francoise Benay, Sarah Kennedy, Kate Semsar, Jenny Knight, Jia Shi, Michelle Smith, Tamara Kelly, Jared Taylor, Harald Yurk, Gulnur Birol, Laurie Langdon, Tom Pentecost, Jackie Stewart, Leilani Arthurs, Andrea Bair, Jennifer Stempien, Brett Gilley, Francis Jones, Ben Kennedy, Stephanie Chasteen, and Beth Simon**, “Clicker Resource Guide: An Instructors Guide to the Effective Use of Personal Response Systems (Clickers) in Teaching,” *Carl Wieman Science Education Initiative, University of British Columbia*, 2017. <<http://www.cwsei.ubc.ca/resources/clickers.htm>>. 8, 12, 19
- Willingham, Daniel T.**, *Why Don’t Students Like School?: A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom*, John Wiley & Sons, 2009. <<http://www.wiley.com/WileyCDA/WileyTitle/productCd-047059196X.html>>. 1, 5, 10, 11, 13, 14