

5-2020

How Often Does Active Learning Actually Occur? Perception versus Reality

Brandon J. Sheridan
Elon University, bsheridan4@elon.edu

Ben O. Smith
University of Nebraska at Omaha, bosmith@unomaha.edu

Follow this and additional works at: <https://digitalcommons.unomaha.edu/econrealestatefacpub>

 Part of the [Economics Commons](#)

Recommended Citation

Sheridan, Brandon J., and Ben Smith. 2020. "How Often Does Active Learning Actually Occur? Perception versus Reality." *AEA Papers and Proceedings*, 110: 304-08.

This Article is brought to you for free and open access by the Department of Economics at DigitalCommons@UNO. It has been accepted for inclusion in Economics Faculty Publications by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.



How Often Does Active Learning Actually Occur? Perception versus Reality[†]

By BRANDON J. SHERIDAN AND BEN SMITH*

I. Measuring Active and Passive Learning

There is now a robust literature touting the benefits of various active learning techniques relative to passive learning pedagogy such as lecturing (e.g., Freeman et al. 2014, Emerson and English 2016, Swoboda and Feiler 2016, Caviglia-Harris 2016). However, recent studies suggest that lecturing is still the dominant pedagogical choice in economics, even though most instructors believe that active learning methods are superior (Goffe and Kauper 2014, Watts and Schaur 2011). A limitation of these studies is that estimates of passive and active learning are based on instructors' subjective, self-reported data.

In contrast, our contribution is to use a new technology, and a well-known survey, to estimate the accuracy of survey measures of active learning. We obtain audio recordings of multiple classes for all instructors in our sample, then match these data to their survey responses. In our sample, instructors overestimate the proportion of time they spend on active learning activities and underestimate the time they spend using passive learning pedagogy. This difference (10.5 percent in mean; 11.5 percent in median) is statistically significant both when treating the data cardinally (t -test, p -value = 0.002) and ordinally (Mann–Whitney (MW) test (Mann and Whitney 1947), p -value = 0.006); the latter test is nonparametric and insensitive to outliers.

* Sheridan: Elon University (email: bsheridan4@elon.edu); Smith: University of Nebraska Omaha (email: bosmith@unomaha.edu). We thank the instructors who allowed us to collect data from their courses; these data have been anonymized here. We also thank KimMarie McGoldrick and Georg Schaur for extremely useful feedback.

[†] Go to <https://doi.org/10.1257/pandp.20201053> to visit the article page for additional materials and author disclosure statement(s).

We appeal to the definition of “active learning” that Freeman et al. (2014, pp. 8413–14) provides in its meta-analysis: “Active learning engages students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert. It emphasizes higher-order thinking and often involves group work.” Common examples include think-pair-share, collaborative learning, and team-based learning. We make use of a new tool to quantitatively and objectively measure active learning that takes place during each class session throughout the semester across a range of business school courses, including many in economics. Owens et al. (2017, p. 3085) developed a software tool in which human classroom observers were used to train an algorithm called Decibel Analysis for Research in Teaching (DART) that can “systematically inventory the presence of active learning with 90 percent accuracy.”¹ This is important because it allows us to capture a continuous measure of how classroom time is used.

Existing data were typically collected via surveys after a course ended—sometimes many semesters afterward, which amplifies reliability concerns. Goffe and Kauper (2014) finds that the mean and median instructor devotes approximately 60 percent of class time to lecturing and 20 percent to instructor-led discussion, each of which DART would classify as passive learning. Watts and Schaur (2011), in a quinquennial survey of faculty, finds that the median instructor spends approximately 83 percent of class time lecturing, with an average value of 65 percent. The authors ask those surveyed to classify their typical lecture time on a 0–4 scale, which corresponds to discrete time blocks of 0, 1–10, 11–33, 34–65, and 66–100 percent,

¹ Note that active learning refers to a pedagogical method, not an outcome.

respectively. They then take the midpoint of these time blocks to construct their descriptive statistics. Obviously, these measures, like all survey data, are imprecise. As in Watts and Schaur (2011), we survey instructors about their usage of class time; we augment these survey measurements with a tool that tracks active and passive learning in all sessions of a given course.

There are several tools developed specifically to describe, in detail, what an instructor actually does during a given class period (e.g., Smith et al. 2013). An advantage of these tools is that the instructor's practices are cataloged as they happen by an independent, trained observer. However, a clear disadvantage is that, in addition to the dozens of hours of training required to learn to use the instrument, someone has to either physically observe or listen to a recording of the instructor's class. This, in itself, is extremely time consuming on the part of the observer and limits the amount of data that can be collected. Moreover, observations from one or two classes of a course are unlikely to be representative of the instructor's broader approach. Therefore, a major contribution of our study is to use DART to analyze audio recordings from entire courses, rather than just one or two classes, to more accurately identify how much time an instructor spends using passive learning strategies relative to nonlecture activities.

II. DART Data and Survey Results

Throughout a given semester, we record classroom audio such that it can be analyzed using the DART software. Further, we survey faculty about their teaching and pedagogical approach in the recorded class. We compare their self-reported responses to the aggregate audio data from their classes. This "perception versus reality" exercise is quite revealing.

A. DART Description

Owens et al. (2017) developed DART to provide a tool that objectively and quickly determines how much active learning occurs during a given class session. The authors essentially trained software to capture what human observers would typically document during a classroom visit, using audio from 1,486 class sessions across 67 different courses. The

software classifies audio recordings into three categories: single voice (S), multiple voice (M), and no voice (N). Based on human classroom observations, they show that time spent in single voice is likely passive learning (e.g., lecturing, instructor-led Q and A), whereas time spent in multiple voice or no voice is most likely a sign of active learning. For example, instructors using clickers usually have a brief period where there is silence as students ponder a question, then many people talking at once when students are discussing answer possibilities with one another. We follow Owens et al. (2017) and classify our recordings to reflect single voice as passive learning and multiple and/or no voice as active learning.

B. DART Data

We collect data from various business school disciplines, with the majority coming from economics. Instructors voluntarily choose to either record their classes on their own or have their classes recorded by lecture capture technology, when available. Recordings begin and end with the official start and end times of the class; recordings are "trimmed" before analysis to minimize the influence of pre- and postclass noise on the recordings.

We then use the DART software to analyze each recording. The output generated by the program includes a chronology of how teaching practices change throughout a given class and the percentage breakdown of how much of each class is spent in single, multiple, and no voice. Collectively, we obtain recordings of 535 class sessions from 30 different instructors.

In Table 1, notice that an average of 89 percent of the time across all 535 class sessions is classified as single voice. As we discuss above, this most likely represents lecturing or, possibly, discussion with one person speaking at a time. Time spent in multiple voice averages slightly more than 9 percent of class time; in theory, this is time when students are actively collaborating or engaging in problem solving. The final column shows the time when the classroom is relatively quiet. This could be a situation in which students are writing out the solution to a problem or participating in the "think" part of a think-pair-share exercise, among other possibilities. Thus, the data show passive learning to be a pervasive reality.

TABLE 1—SHARE OF CLASS TIME WITH SINGLE, MULTIPLE, AND NO VOICES AUDIBLE

	Single	Multiple	None
Average percentage (SD)	89.0 (7.2)	9.4 (6.3)	1.6 (2.3)
Median percentage	90.1	7.9	0.6
Total class sessions			535

Note: Data generated from DART analysis of classroom recordings.

C. Survey Results

We also ask instructors to complete a survey about their attitudes and teaching practices. We adopt a subset of the questions from Watts and Schaur (2011) to allow us to compare instructors’ perceptions of their teaching with the reality of their DART data. Instructors are relatively evenly distributed as tenured, tenure-track without tenure, and non-tenure-track; the median instructor has seven years of experience. The primary methods of instruction are all some form of the instructor delivering content and students passively receiving it.

Table 2 shows the most salient results. Simply put, instructors greatly underestimate how much they lecture or otherwise use passive learning techniques. Recall that the actual average time spent in single voice in Table 1 is 89 percent. In our raw survey data (not reported here), instructors estimate that they spend only about 65 percent of their time lecturing. We combine this with reported estimates of time spent on instructor-led discussion and videos to complete our estimated measure of passive learning, and to be consistent with the way DART codes the audio output; this leaves us with a perception of 78.5 percent. The sample average reality-perception gap is thus 10.5 percentage points; the central tendency of the perception distribution differs from reality at the 1 percent significance level using both a simple *t*-test (*p*-value = 0.002) and an MW test (*p*-value = 0.006). To maintain the independence assumption, we aggregate our DART data to the course level for all statistical tests.² Further, this result persists and remains consistent when

²It is of note that instructors are asked for averages across the course on the survey instrument. In principle, we are aggregating the DART data to the same level as the survey instrument (the course). There is within-course variation that

TABLE 2—OBJECTIVE VERSUS SUBJECTIVE MEASURES OF INSTRUCTION

Learning type	Time spent (percentage)		
	Reality	Perception	Gap
Passive (average)	89.0	78.5	10.5
Passive (median)	90.1	80.0	11.1
Active (average)	11.0	21.5	−10.5
Active (median)	8.5	20.0	−11.5

Note: The reality column is the DART data, while the perception column is the instructor survey data.

we look at instructor-specific gaps, for which the average is 8.9 percent (median 9.1 percent) more passive learning than perceived.

As rank-based tests are not commonly used in economics education, we thought it might be helpful to explain the inner workings of the MW test so the reader can better interpret the results. To determine whether there is a statistically significant difference in the central tendency of the two samples, the data from both samples are combined into a single set and ordered from lowest to highest. In essence, this test is summing the number of observations in sample two that have a greater rank than each of the observations in sample one. This produces the *U*-test statistic; if this statistic is greater than the critical value generated from the distribution of n_1, n_2 elements randomly ordered, the null hypothesis is rejected. The two notable properties of this test are that (i) there is no underlying distribution assumption and (ii) outliers have a minimal impact because all results are converted to ranks. Given the advantages of this statistical test, we are confident that the instructors in our dataset use passive learning more than they report.

This underestimate of passive learning does not appear to be driven by a subgroup of the data (e.g., mean/median values of tenure-track versus non-tenure-track faculty are not substantively different). However, the subgroups are too small to conduct statistical tests, and our ability to detect differences between these groups is therefore limited, so we do not report those results here. The kernel density estimates (KDEs; Figure 1, panel A) show that the mass

we are averaging out with this approach. However, our intent is to match the survey instrument as closely as possible.

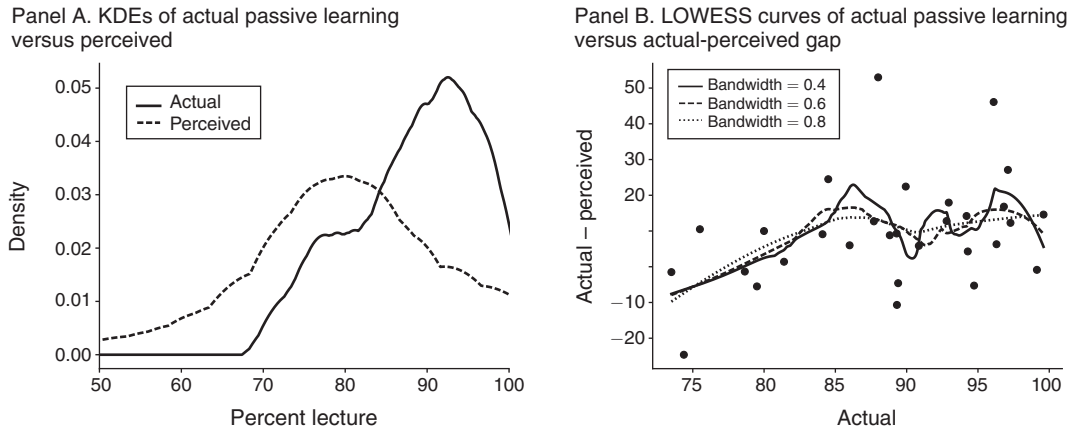


FIGURE 1. KDE AND LOWESS CURVES COMPARING ACTUAL VERSUS PERCEIVED PASSIVE LEARNING

of the actual passive learning distribution is consistently higher (less variation) than the perceived passive learning distribution. However, we do see a systematic difference in instructor perception error based on the amount of passive learning; as it increases, instructors increasingly underestimate the passive learning in their classroom (Figure 1, panel B). Thus, while instructors seem to understand that they lecture a lot, they still consistently underestimate how much they actually lecture. It is notable that locally weighted scatterplot smoothing (LOWESS) curves are not very sensitive to outliers, as they use subsets of the data; no single data point can drive the entire curve. Note that each dot represents one instructor matching the instructor's survey response, and regardless of the chosen bandwidth, there is a general upward trend.

III. Discussion and Future Research

The effectiveness of active learning in the economics classroom, among other disciplines, is widely taken as given among instructors. Paradoxically, self-reported survey measures reveal that passive learning (e.g., lecturing) is the dominant pedagogy. We improve upon previous measures by using continuous, objective data to show that not only do instructors lecture a lot but they lecture a lot more than they think they do. This gap between perception and reality may occur for many reasons. For example, for instructors who predominantly lecture, using active learning strategies may prove

challenging, and research shows that performing a challenging task distorts our perception of time, making it appear to go slower (Eagleman 2008). Thus, instructors mistakenly overestimate the amount of time they spend using active learning techniques. Alternatively, instructors might misremember the most salient part of class as engaging and active, or because active learning is considered a "good" form of teaching, it may be psychologically easier to believe that it accounts for a relatively high proportion of class time. This warrants further investigation.

We caution instructors considering using DART that it is a tool and is only as good as the user's understanding of its strengths and weaknesses (for details, see Owens et al. 2017). For example, the tool cannot yet distinguish among different voices, so analyzing engagement during classroom discussions is challenging. The tool does an excellent job of cataloging which times during a class are most likely to consist of active learning. However, the tool cannot measure how effectively one uses a particular active learning technique. For most instructors, evidence of good teaching is vital for job security. This tool allows them to see where they are and to track their progress over time in a concrete way.

REFERENCES

- Caviglia-Harris, Jill. 2016. "Flipping the Undergraduate Economics Classroom: Using Online Videos to Enhance Teaching and Learning." *Southern Economic Journal* 83 (1): 321–31.

- Eagleman, David M.** 2008. "Human Time Perception and Its Illusions." *Current Opinion in Neurobiology* 18 (2): 131–36.
- Emerson, Tisha L. N., and Linda English.** 2016. "Classroom Experiments: Is More More?" *American Economic Review* 106 (5): 363–67.
- Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth.** 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics." *Proceedings of the National Academy of Sciences of the United States of America* 111 (23): 8410–15.
- Goffe, William L., and David Kauper.** 2014. "A Survey of Principles Instructors: Why Lecture Prevails." *Journal of Economic Education* 45 (4): 360–75.
- Mann, H. B., and D. R. Whitney.** 1947. "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other." *Annals of Mathematical Statistics* 18 (1): 50–60.
- Owens, Melinda T., Shannon B. Seidel, Mike Wong, Travis E. Bejines, Susanne Lietz, Joseph R. Perez, Shangheng Sit, et al.** 2017. "Classroom Sound Can Be Used to Classify Teaching Practices in College Science Courses." *Proceedings of the National Academy of Sciences of the United States of America* 114 (12): 3085–90.
- Smith, Michelle K., Francis H. M. Jones, Sarah L. Gilbert, and Carl E. Wieman.** 2013. "The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices." *CBE—Life Sciences Education* 12 (4): 618–27.
- Swoboda, Aaron, and Lauren Feiler.** 2016. "Measuring the Effect of Blended Learning: Evidence from a Selective Liberal Arts College." *American Economic Review* 106 (5): 368–72.
- Watts, Michael, and Georg Schaur.** 2011. "Teaching and Assessment Methods in Undergraduate Economics: A Fourth National Quinquennial Survey." *Journal of Economic Education* 42 (3): 294–309.