How often does active learning occur? A quantitative approach.

By Brandon J. Sheridan and Ben Smith*

I. Active Learning and the Persistence of Lecturing

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 lecturing and active learning are based on instructors' subjective, self-reported data. In contrast, we use a new technology, and a well-known survey, to estimate the accuracy of survey measures of active learning. In our sample, instructors overestimate the proportion of the time they spend on active learning activities and underestimate the time they spend lecturing. This difference (10.5% in mean; 11.5% in median) is statistically significantly different both when treating the data cardinally (t-test, p-value = 0.002) or as ordinal (Mann and Whitney, 1947, p-value = 0.006); the latter is a non-parametric test that is insensitive to outliers.

We appeal to the definition of 'active learning' Freeman et al. (2014) provide in their 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."[1] 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) developed a software tool in which they used human classroom observers to train an algorithm they call *Decibel Analysis for Research in Teaching* (henceforth DART) that can "systematically inventory the presence of active learning with 90% accuracy."[2] 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 afterwards, which amplifies reliability concerns. Goffe and Kauper (2014) find that the mean and median instructor devotes approximately 60% of time to lecturing and 20% to instructor-led discussion, each of which DART would classify as passive learning. Watts and Schaur (2011), in their quinquennial survey of faculty, find that the median instructor spends approximately 83% of class time lecturing with an average value of 65%. They 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, 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.

Goffe and Kauper (2014) address the question of lecture persistence in their sur-

* Sheridan: Elon University, 2075 Campus Box, Elon, NC 27302, bsheridan4@elon.edu. Smith: University of Nebraska at Omaha, 6001 Dodge Street, Omaha, NE 68182, bosmith@unomaha.edu.

[1] Freeman et al. (2014), pgs. 8413-8414.

[2] Owens et al. (2017), pg. 3085. Note that active learning refers to a pedagogical method, not an outcome.

vey of 340 economics principles instructors. They find that two-thirds believe lecturing is not the best teaching method but, of those, over 40% think lecturing is cost-effective. That is, even if active learning methods may improve outcomes for students, some instructors think the time cost involved in implementing such approaches outweighs the resulting marginal success of students. Further, it can be difficult for an instructor to tell whether students perform meaningfully better in a classroom that uses more active learning methods. It is also possible that an inexperienced instructor, or an experienced instructor using active learning strategies for the first time, may not properly use the techniques, leading to lackluster enthusiasm and/or performance on the part of students (or instructor). Consequently, the instructor may return to their previous method of instruction which, often, is lecturing.

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 the instructor's practices are catalogued as they happen by an independent, trained observer. However, a clear disadvantage is in addition to dozens of hours of training time required to learn to use the instrument, someone has to either physically observe or watch/listen 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 lecturing or in other passive learning formats relative to non-lecture 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 audio data from their classes. This "perceptions vs. 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. They 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 & 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. (2018) and classify our recordings to reflect single voice as lecturing and multiple and/or no voices as active learning.

B. DART Data

We collect data from various business-school disciplines, with the majority being from economics. Instructors voluntarily chose to either record their classes on their own or have their classes recorded by lecture capture technology, when available. Recordings began and ended with the official start and end times of the class; recordings were 'trimmed' before analysis to minimize pre/post-class noise from influencing the recordings.

We then used 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 obtained recordings of 535 class sessions from 30 different instructors.

	Single	Multiple	None
Avg. Percentage	89.0	9.4	1.6
(Std. Dev.)	(7.2)	(6.3)	(2.3)
Med. Percentage	90.1	7.9	0.6
Total Class Sessions			535

Table 1—: DART Summary

In Table 1, notice an average of 89% of the time across all 535 class sessions was classified as single voice. As we discussed earlier, 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% 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, participating in the "think" part of a think-pair-share exercise, among other possibilities. Thus, data show passive learning to be a pervasive reality.

C. *Survey Results*

We also asked 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 7 years of experience. The primary methods of instruction are all some form of the instructor delivering content and students passively receiving it.

Learning Type	Time Spent: Reality	Perception	Gap
Passive (Avg.)	89.0%	78.5%	10.5%
Passive (Med.)	90.1%	80.0%	11.1%
Active (Avg.)	11.0%	21.5%	-10.5%
Active (Med.)	8.5%	20.0%	-11.5%
Decreasing Returns occur in ≤ ___ mins:			
	20	30	40
Passive	48%	26%	26%
Active	72%	17%	11%

Table 2—: Faculty Perceptions of Teaching Survey

Table 2 shows the most salient results. Simply put, instructors greatly underestimate how much they lecture or otherwise use passive learning techniques. Recall the actual average time spent in single voice in Table 1 was 89%, whereas instructors estimate they spend only about 65% of their time lecturing. We also include discussion and videos as a form of passive learning to be consistent with the way DART codes the audio output; this still leaves us with a perception of 78.5%. The sample average reality-perception gap is thus 10.5%; the central tendency of the perception distribution differs from reality at the 1% significance level using both a simple t-test (p-value = 0.002) and a Mann-Whitney (Mann and Whitney, 1947) test (p-value = 0.006). Further, this result persists, and remains consistent, when we look at instructor-specific gaps, for which the average is 8.9% (median 9.1%) more passive learning than perceived. We also find it striking that instructors think decreasing returns occur much quicker to active learning. That is, instructors seem to be much more confident in a student's ability to learn by listening rather than learn by actively engaging with the material during a given class.

As rank-based tests are not commonly used in Economic Education, the authors thought it might be helpful to explain the inner workings of the Mann-Whitney (MW) test so the reader can better interpret the results. To determine if 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. A column is added to the data indicating the 'rank' of the observation (ordered from 1 to $n_1 + n_2$); the MW test uses this rank data to perform the statistical test. The sum of the ranks for a given sample is then calculated and adjusted for sample size. 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 statistic. The comparison distribution is generated from the assumption that if the two groups were identical their order would be a random shuffle from all possible rank orders. If the U statistic is greater than the critical value generated from the distribution of n_1, n_2 elements randomly ordered, the null hypothesis can be rejected. The two notable properties of this test are: (1) there is no underlying distribution assumption and (2) outliers have a minimal impact on the U statistic as all results are converted to ranks. Given the advantages of this statistical test, we are confident that the instructors in our dataset lecture more than they report.

The overestimate of active learning does not appear to be driven by a subgroup of the data (e.g. mean/median values of tenure vs. non-tenure track faculty are not substantively different); though, the subgroups are quite small, thus our ability to detect differences between these groups is limited. The Kernel Density Estimates (Figure 1a) show that the mass of the actual lecture distribution is consistently higher than the perceived lecture distribution. However, we do see a systematic difference in instructor perception error based on the amount of lecture. As the amount of lecture increases, instructors increasingly overestimate the active learning in their classroom (Figure 1b). Thus, while instructors seem to understand they lecture a lot, they still consistently underestimate how much they actually lecture. It is notable that LOWESS curves (locally weighted scatterplot smoothing) are not very sensitive to outliers as they use subsets of the data; no single data point can drive the entire curve.

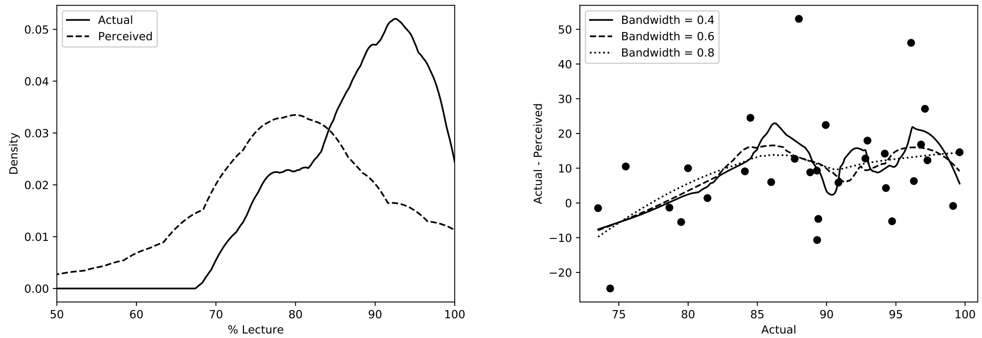
III. Discussion & Future Research

The effectiveness of active learning in the economics classroom is widely taken as given among instructors. Paradoxically, self-reported survey measures reveal that passive learning (e.g. lecturing) is more common. We improve upon previous measures by using continuous, objective data to show that not only do instructors lecture a lot, 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 mis-remember the most salient part of class as engaging and active, or because active learning is considered a 'good' form of teaching, it could make it psychologically easier to believe 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 (see Owens et al. (2018) for details). For example, the tool cannot yet distinguish between different voices, so analyzing engagement during classroom discussions is challenging. The tool does an excellent job of cataloguing which times during a class are most likely to consist of active learning. However, the tool cannot measure how effectively you use 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 track their progress over time in a concrete way.

REFERENCES

Caviglia-Harris, J. (2016). Flipping the undergraduate economics classroom: Using online videos to enhance teaching



(a) KDEs of actual vs perceived

(b) LOWESS curves of actual vs actual-perceived

Figure 1. : KDE and LOWESS curves comparing actual lecture vs perceived lecture. The three LOWESS curves are fit to this data at three different bandwidths. Regardless of the curve, there is a general upward trend where faculty who lecture a very high percentage of the time overestimate the time spent on active learning more.

and learning. *Southern Economic Journal 83*(1), 321–331.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current opinion in neurobiology 18*(2), 131–136.

Emerson, T. L. and L. English (2016). Classroom experiments: Is more more? *American Economic Review 106*(5), 363–67.

Freeman, S., S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences 111*(23), 8410–8415.

Goffe, W. L. and D. Kauper (2014). A survey of principles instructors: Why lecture prevails. *The Journal of Economic Education 45*(4), 360–375.

Mann, H. B. and D. R. Whitney (1947, 03). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist. 18*(1), 50–60.

Owens, M. T., S. B. Seidel, M. Wong, T. E. Bejines, S. Lietz, J. R. Perez, S. Sit, Z.-S. Subedar, G. N. Acker, S. F. Akana, et al. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences 114*(12), 3085–3090.

Owens, M. T., G. Trujillo, S. B. Seidel, C. D. Harrison, K. M. Farrar, H. P. Benton, J. Blair, K. E. Boyer, J. L. Breckler, L. W. Burrus, et al. (2018). Collectively improving our teaching: attempting biology department–wide professional development in scientific teaching. *CBE—Life Sciences Education 17*(1), ar2.

Smith, M. K., F. H. Jones, S. L. Gilbert, and C. 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), 618–627.

Swoboda, A. and L. Feiler (2016). Measuring the effect of blended learning: Evidence from a selective liberal arts college. *American Economic Review 106*(5), 368–72.

Watts, M. and G. Schaur (2011). Teaching and assessment methods in undergraduate economics: A fourth national quinquennial survey. *The Journal of Economic Education 42*(3), 294–309.