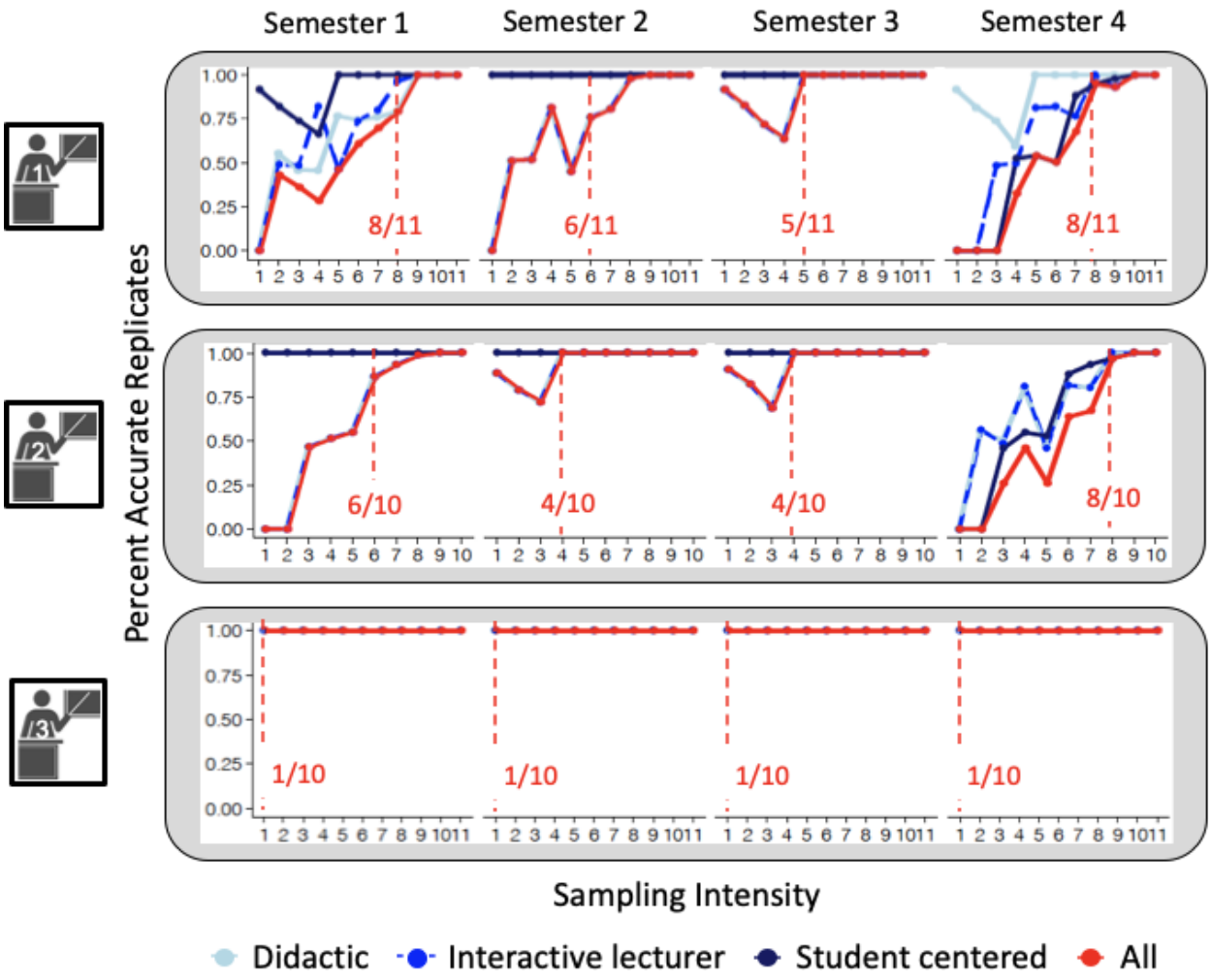
Supplemental Material

CBE—Life Sciences Education
Sbeglia *et al*.

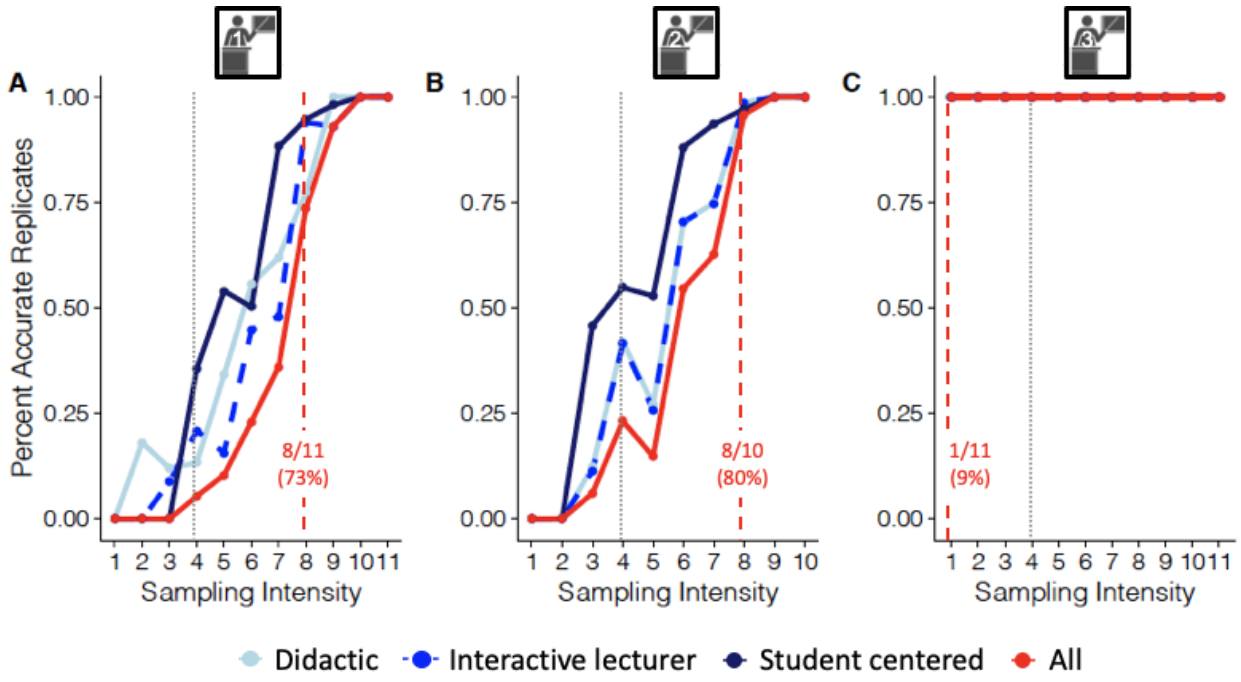
Supplemental Table 1. The 25 observational codes and behaviors of the COPUS.

	Code	Behavioral Code Description
Instructor	1o1*	One-on-one extended discussion with one or a few individuals
	Adm	Administration (assign homework, return tests, etc.).
	CQ*	Asking a clicker question
	D/V	Showing or conducting a demo, experiment, simulation or video.
	Fup	Follow-up/feedback on clicker question or activity to entire class.
	LAnQ	Listening to and answering student questions with entire class listening.
	Lec*	Lecturing (presenting content, presenting a problem solution, etc.)
	LO	Other - explain in comments
	LW	Waiting (not interacting or observing/listening to student activity in class)
	MG	Moving through class guiding ongoing student work
	PQ*	Posing a non-clicker question to students (non-rhetorical).
	RtW	Real-time writing on board, doc. projector, etc.
Student	CG*	Discuss clicker question in groups of 2 or more students.
	Ind	Individual thinking/problem solving
	L	Listening to instructor/taking notes, etc.
	OG*	Other assigned group activity, such as responding to instructor question.
	Prd	Making a prediction about the outcome of a demo or experiment.
	SAnQ*	Student answering a question posed by the instructor
	SO	Other - explain in comments.
	SP	Presentation by student(s)
	SQ	Student asks question.
	SW	Waiting (instructor late, working on fixing AV problems etc.)
	TQ	Test or quiz
	WC	Engaged in whole class discussion
	WG*	Working in groups on worksheet activity.

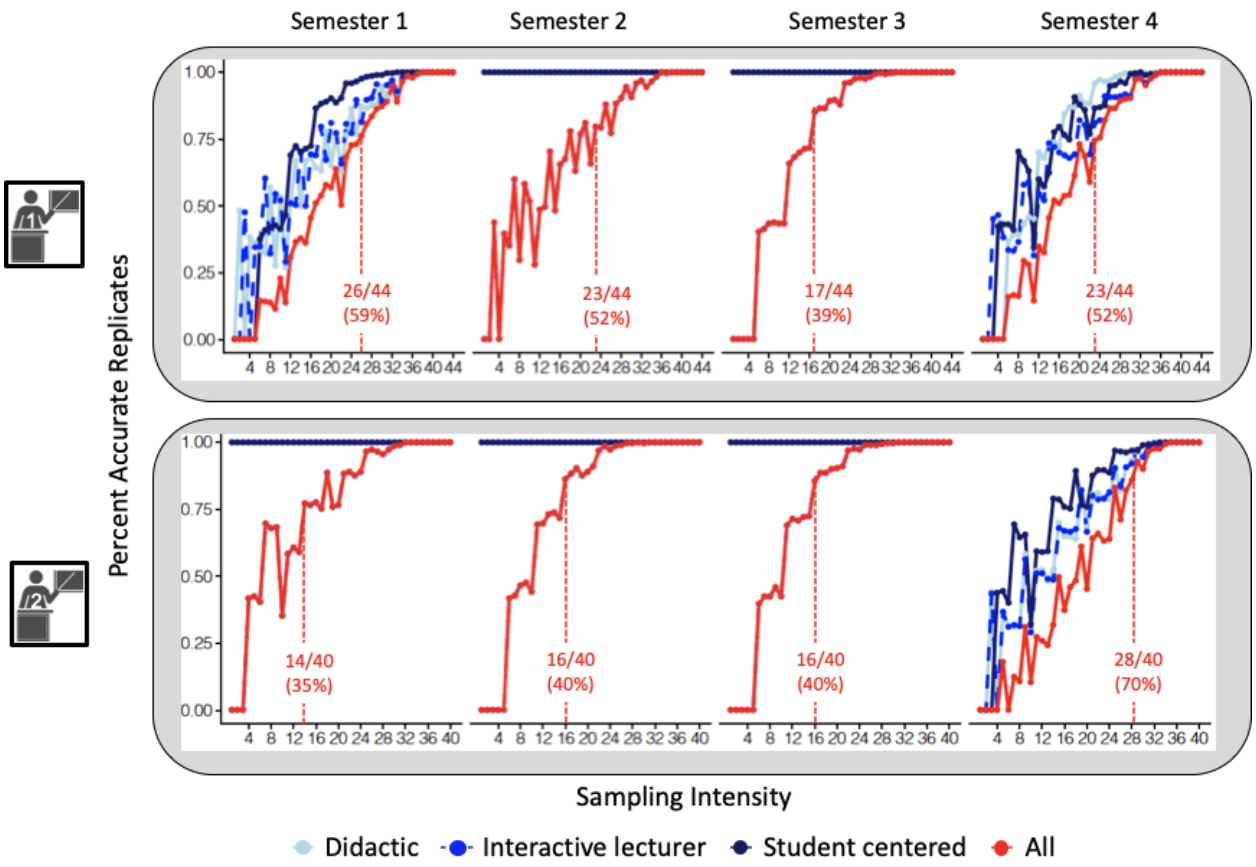
*Behaviors used by the COPUSAnalyzer



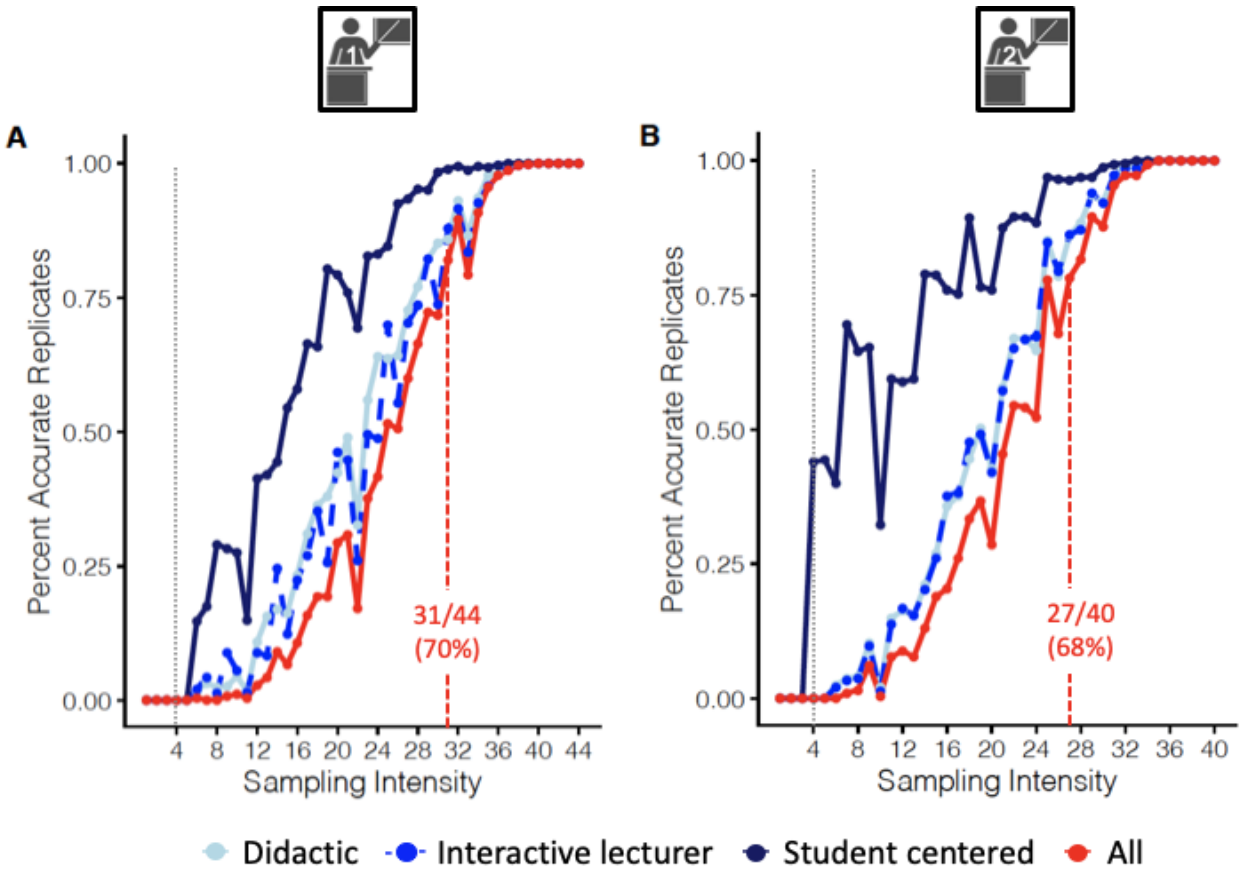
Supplemental Figure 1. Simulation results for each instructor and semester in our dataset using a 15% threshold for accurate estimation. The sampling intensity required to attain a ≥ 75% probability of accurately estimating all instructional styles in each semester is indicated by the vertical red dashed line. Sampling intensities with accuracy probabilities above this 75% threshold can be considered capable of accurate and precise measurement for the instructors and courses in this sample. Note that these results are based on 128 observations (full classes) gathered for this study. At this threshold, the number of classes needed to accurately estimate the proportion of instructional styles within these courses ranged from 1-8.



Supplemental Figure 2. Simulation results through time for instructors 1 (A), 2 (B), and 3 (C) using a 15% threshold of accurate estimation. The sampling intensity required to attain a ≥ 75% probability of accurately estimating all instructional styles in every semester sampled for each instructor is indicated by the red dashed line and text. Sampling intensities with accuracy probabilities above this 75% threshold can be considered capable of accurate and precise measurement for the instructors and courses in this sample. For Instructor 3, whose classes were consistently didactic, a sampling intensity of one class was sufficient to accurately measure change (or lack thereof) in the classroom learning environment through time. Conversely, the courses for Instructors 1 and 2 were characterized by two instructional styles some semesters and three in others, and the proportions shifted through time. Given the inconsistency of their styles through time, higher sampling intensities (i.e., 8 classes) were required to measure change in these instructors through time with a high probability.



Supplemental Figure 3. Simulation results for each instructor and semester in our dataset using a 9% threshold for accurate estimation. The data were copied four times to simulate a course with 40-44 classes. Instructor 3 was not included in this analysis because they showed no variation in their instructional styles. The sampling intensity required to attain a ≥ 75% probability of accurately estimating all instructional styles in each semester is indicated by the vertical red dashed line. Sampling intensities with accuracy probabilities above this 75% threshold can be considered capable of accurate and precise measurement for the instructors and courses in this sample. Note that these results are based on 128 observations (full classes) gathered for this study. At this threshold, the number of classes needed to accurately estimate the proportion of instructional styles within these courses ranged from 14-28 (35%-70%).



Supplemental Figure 4. Simulation results through time for instructors 1 (A) and 2 (B) using a 9% threshold of accurate estimation. The data were copied four times to simulate a course with 40-44 classes. Instructor 3 was not included in this analysis because they showed no variation in their instructional styles. The sampling intensity required to attain a $\geq$ 75% probability of accurately estimating all instructional styles in every semester sampled for each instructor is indicated by the red dashed line and text. Sampling intensities with accuracy probabilities above this 75% threshold can be considered capable of accurate and precise measurement for the instructors and courses in this sample. For Instructor 3, whose classes were consistently didactic, a sampling intensity of one class was sufficient to accurately measure change (or lack thereof) in the classroom learning environment through time. Conversely, the courses for Instructors 1 and 2 were characterized by two instructional styles some semesters and three in others, and the proportions shifted through time. Given the inconsistency of their styles through time, higher sampling intensities (i.e., eight classes) were required to measure change in these instructors through time with a high probability.