Characterizing active learning environments in physics using latent profile analysis

Kelley Commeford and Eric Brewe
Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104, USA

Adrienne Traxler
Wright State University, 3640 Colonel Glenn Highway, Dayton, Ohio 45435, USA

(Received 10 May 2021; accepted 5 January 2022; published 23 February 2022)

The vast majority of research involving active learning pedagogies uses passive lecture methods as a baseline. We propose to move beyond such comparisons to understand the mechanisms that make different active learning styles unique. Here, we use COPUS observations to record student and instructor activities in six known styles of active learning in physics, and use latent profile analysis to classify these observations. Latent profile analysis using two profiles successfully groups COPUS profiles into interactive lecturelike and other. Five latent profiles successfully sorts observations into interactive lecturelike, Modeling Instruction, ISLE labs, context-rich problems labs, and recitationlike or discussionlike. This analysis serves as a proof of concept, and suggests instructional differences across pedagogies that can be further investigated using this method.

DOI: 10.1103/PhysRevPhysEducRes.18.010113

I. INTRODUCTION

Active learning has been gaining traction in physics classrooms, with numerous studies showing how active learning can promote learning gains over passive lecture alternatives. However, within the umbrella of active learning, several physics pedagogies have spawned from different ideological foundations, often with wildly varying mechanical differences [1–3]. As such, the term "active learning" no longer sufficiently describes a pedagogy, making it difficult to pinpoint the mechanisms that lead to observed learning gains. Thus, it is imperative that we understand the mechanisms that make different "active learning in physics" pedagogies unique, so future studies can further delineate the benefits of specific active learning mechanisms.

The "characterizing active learning environments in physics" project, or CALEP, sets out to establish a vocabulary that will allow us to speak about the different active learning pedagogies in physics, without relying on a comparison to passive lecture methods. We focused on two aspects of the active learning pedagogies; how students and instructors spend time in class, and the student networks that result from the instruction. A full summary of the CALEP project can be found in Ref. [4]. We collected classroom observations and student network data from six prominent research-based introductory physics curricula. We used the classroom observation protocol for undergraduate science, technology, engineering, and mathematics (STEM) courses. This protocol, abbreviated as COPUS [5], allows us to measure the time dedicated to certain activities during a class period, which, in essence, gives us an idea of the kinds of activities that occur in different active learning environments. For network analysis results, please refer to Refs. [4,6].

This paper will focus on the COPUS observations, where we discuss the results of latent profile analysis (LPA) on the observed COPUS profiles. LPA can be used to tease out hidden categories within data—such as physics pedagogies. LPA has been used with COPUS data to describe a large sample of STEM classes, which effectively sorted them into active, passive, and hybrid categories [7]. Our analysis builds on that work by restricting to known active learning pedagogies in physics classrooms, as a first effort at developing a more nuanced classification scheme.

II. BACKGROUND

Because active learning is an umbrella term for different pedagogical approaches, it fails to identify distinguishing characteristics of pedagogies in physics. To this end, we identified six pedagogies that are well represented in the literature and at professional development workshops at national conferences. These six pedagogies have different approaches to active learning. In order to explore these different approaches, we used COPUS as a means to understanding unique features of classroom activity.

TABLE I. COPUS observation example. Time is measured in two-minute intervals.

	Students doing													Instructor doing										
Time	L	IND	CG	WG	OG	AnQ	SQ	WC	Prd	SP	T/Q	W	O	Lec	RtW	Fup	PQ	CQ	AnQ	MG	1o1	D/V	Adm	W	O
0–2				×									×							×	×			×	
2–4				×			×						×							×	×				
4–6				×			×						×							×	×				
6–8				×									×							×	×				
8–10				×									×							×	×				
10–12				×			×													×	×				
12–14				×																	×				
14–16				×			×													×	×				
16–18				×																×	×			×	
18–20				×																×	×				
Total	0	0	0	10	0	0	4	0	0	0	0	0	5	0	0	0	0	0	0	9	10	0	0	2	0

A. Active learning pedagogies in physics

We collected data from six well-known active learning pedagogies in physics. The pedagogies we studied were Tutorials in Introductory Physics, the Investigative Science Learning Environment (ISLE), Modeling Instruction, Peer Instruction, context-rich problems (aka the Minnesota curriculum), and SCALE-UP. We reached out directly to the developers of the curricula to identify high-fidelity implementations of each pedagogy. When possible, the institution that developed the pedagogy was used as a research site; however, this was not always feasible. Secondary institutions were identified via recommendation of instructors with extensive training or research experience with the pedagogy in question.

Tutorials in Introductory Physics maintains a lecture-lab-recitation structure typical of large introductory courses. The recitations, or tutorial sections, are interactive small group environments that have the students follow scaffolded worksheets that emphasize misconception confrontation and resolution. The teaching assistants facilitate small group interactions [8,9].

ISLE treats students as novice scientists, and focuses on building up correct intuition rather than debunking misconceptions [10,11]. Students work in small groups through a lab-based learning cycle, while the instructors facilitate discussion.

Modeling Instruction is a fully contained learning environment where students work in small groups to develop, test, deploy, and revise conceptual models, and then come together for large "white-board meetings" to discuss their findings [12–14]. The instructors facilitate small group discussion and guide the larger white-board meetings.

Peer Instruction is a structured, interactive lecture environment, typically used when classroom constraints deem a large lecture hall necessary. Peer Instruction divides a class session into small modules, which begin with a short lecture, and cycle through several clicker questions. Students answer individually and then re-answer after discussing with their peers [15,16].

Context-rich problems leads students through the course under a pseudo apprentice-expert relationship. This pedagogy uses context-rich problems in all aspects of the course, to teach students how to identify relevant information and engage in expertlike problem-solving behavior, as demonstrated by the instructor and teaching assistants [17,18].

Finally, SCALE-UP is also a fully contained classroom environment, where the students are in pairs or groups of three, but seated at large tables with other groups of two or three students to facilitate larger discussions. SCALE-UP usually refers more to the classroom setup than the pedagogy itself, but lends itself nicely to the adoption of many active learning styles [19]. As such, it is often treated as an independent pedagogy.

B. The classroom observation protocol for undergraduate STEM

The Classroom Observation Protocol for Undergraduate STEM (COPUS) was developed by Smith *et al.* [5]. It consists of 12 instructor and 13 student codes, for which the observer marks whether the coded behavior occurred or not during two-minute intervals. An activity is counted if the behavior occurs for at least five seconds during the two-minute interval. A full list of COPUS code abbreviations and their descriptions has been included in the Supplemental Material [20]. An example of a COPUS observation for a tutorial session can be seen in Table I.

From these observations, we can compile the selected codes into COPUS profiles. The method of COPUS profile creation is not always clearly reported in literature that uses COPUS, making it difficult to compare analyses. It is

TABLE II. COPUS profile example. The total number of occurrences for each code was tallied in each column, and divided by the number of intervals. For this example, there were 10 intervals.

Students doing		Instructor doing	
L	0	Lec	0
IND	0	RtW	0
CG	0	Fup	0
WG	1.0	PQ	0
OG	0	CQ	0
AnQ	0	AnQ	0
SQ	0.4	MG	0.9
WC	0	1o1	1.0
Prd	0	D/V	0
SP	0	Adm	0
T/Q	0	W	0.2
W	0	O	0
O	0.5		

typical of such studies to report percentages for each code, without a description of how ratios were taken or visualizations to infer the same information. As such, we explain our process here. For the CALEP project, we used the "bar chart" method, in which COPUS profiles were created by summing the number of marks in each column (indicated in the "total" row in Table I), and dividing by the number of intervals over which the observation occurred (10 intervals, using the same example). This method leads to the percentage of two-minute intervals in which a code was present. An example of a COPUS profile from CALEP can be seen in Table II. It should be noted, however, that we calculate our COPUS profiles differently than in Smith *et al.* [5]. In Smith *et al.*, the number of marks is tallied for each column, and divided by the total number of marks across all columns. This method, which we dub the "pie chart" method, is useful when investigating the prevalence of activities. We chose the former method of profile creation, as we felt it provided a more explicit picture of how class time was spent. If an instructor spends the majority of the time lecturing, but peppers in several other instructional methods, the latter method of profile creation can be misleading and underreport the frequency of a code. We wanted a profile that showed the fraction of class time engaged in a certain instructional method, not fraction of every code reported.

C. Latent profile analysis

LPA, sometimes referred to as finite Gaussian mixture modeling, allows us to uncover hidden groupings within a dataset. It allows us to make predictions about unobserved discrete variables based on clustering that occurs within observed variables. LPA can be used for any situation that calls for fitting continuous observed variables into discrete latent variables [21]. For a visualization of how this works, Fig. 1 illustrates some measured variable (in red). It is clear to us that this is not a normal distribution on its own, but could probabilistically be separated into two groups that overlap (blue and green). LPA helps us determine those hidden groups.

LPA is similar to factor analysis in that it uses observed data to identify hidden groupings. Factor analysis uses correlations between variables to identify groupings of variables that are similar, while LPA differs by attempting to split seemingly homogeneous data into discrete groups created by similar variable structure.

For the CALEP project, our goal is to determine what the quantifiable identifying features of different physics pedagogies are. Active learning comes with an abundance of student-centered activities, which, in theory, are different for each pedagogy. As such, we chose to use LPA to see if this analysis method could identify different pedagogies (our unobserved, discrete latent variables, or groups or profiles) based on observations of active learning behaviors (our measured variables). Using this same data structure to further differentiate LPA versus factor analysis, factor analysis would tell us what active learning behaviors are commonly seen together (groupings of variables), while LPA tells us groups of observations that have similar active learning behaviors. While these two methods are very similar, the output of LPA more closely aligns with our goals of identifying hidden pedogogical categories.

We used the percentage of class time per COPUS code (known as the COPUS profile) as our input variables to LPA, with the expectation that LPA would in turn group observations with similar COPUS profiles. Ideally, these groupings would align with the identified active learning pedagogies each observation represents. Unfortunately, our data size was too small to run LPA with variable correlations or factor analysis, so we were unable to compare the different methods.

D. Previous studies using LPA

Stains *et al.* previously utilized LPA with COPUS observations from several science disciplines [7]. They collected COPUS observations from classes across STEM disciplines, most of which were at doctoral-granting institutions, using a convenience approach—a call was put out on discipline-based education research (DBER) listservs asking researchers to share COPUS data, and they also collected voluntary COPUS submissions using the COPUS analyzer website [22]. After using LPA on these solicited COPUS profiles, Stains *et al.* ultimately chose the model with seven profiles for their data, which they further sorted into didactic (mostly lecture), interactive lecture (hybrid), or student centered (active) learning environments.

Stains *et al.* [7] took random observations and *then* classified them into active, passive, or hybrid pedagogies. We differ from Stains *et al.* in that we used purposeful

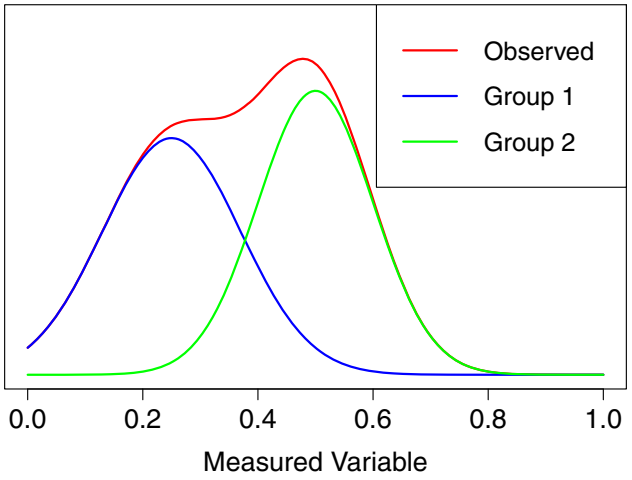

FIG. 1. Given an observed distribution (red), LPA assumes that this observed distribution can be described by a combination of distinct profiles that follow a Gaussian distribution (two groups, shown in blue and green).

sampling, collecting data at sites chosen due to their high fidelity implementation of different active learning pedagogies. Further, we feel that it is no longer sufficient to be abstract in classifying pedagogies as "active" or "not active," so our purpose was to extend the work of Stains *et al.* by zooming in on the active learning category, while limiting our scope to investigate differences among active learning specifically in physics.

We have provided a visual in Fig. 1 to help illustrate how LPA classifies COPUS profiles into distinct profiles. Given a distribution of an observed variable (in our case a single category from COPUS), shown in red, LPA assumes that this observed distribution can be described by a combination of distinct profiles that follow a Gaussian distribution (two groups, shown in blue and green). LPA begins by taking some initial Gaussian distribution to represent each profile, of which you can have any number, then iterates through the data and adjusts the Gaussians until all points are assigned to a profile such that the expectation value is maximized. LPA does this by calculating a probability distribution for each profile (the group Gaussians), and assigns a weighted profile membership to each data point (via probability distribution). It then recalculates the probability distribution representing each profile using the weighted members, and repeats this process until it converges [23].

LPA can also accept constraints to these Gaussians, via variance and covariance. The variance can be equal between profiles (the Gaussians have the same width), or vary between profiles. Additionally, the prevalence of input variables can depend on each other, introducing covariance, which can be set to zero, equal, or varying between profiles in the model as well [24]. Understanding the variance of the profile distributions is a bit more straightforward if we preemptively reference the top panel of Fig. 3; the variance is the size of the box, and equal variance forces the size of the boxes for each code to be equal for each profile. Varying variance, on the other hand, allows these boxes to be different sizes. The covariance describes how the input variables depend on each other, which is a bit harder to visualize. In the context of our data, covariance would tell us how the presence of lecture depends on the presence of real-time-writing. While covariance is likely present in our data (by observation, real-time-writing does typically accompany lecture), we did not have a large enough dataset to get the varying covariance model to converge. Additionally, the equal covariance model, which forces each pedagogy to follow the same relationships between codes, was excluded as it was not an appropriate constraint for our study. Thus, we assume zero covariance in our analysis. Future work should explore the role that covariance plays in the development of profiles, where a larger dataset can illuminate meaningful relationships between COPUS codes.

III. METHODS

We collected COPUS observations from six high fidelity active learning pedagogies in physics [4]. We traveled to the sites and observed as many sections of the course as possible in a one week visitation period. All observations were done in a live environment, in person, by the same observer. The official COPUS recording spreadsheet was used for these data collections [5]. A summary of the number of sections and the number of observations per section can be seen in Table III.

From the base COPUS observations with check marks in each code column, we created the COPUS profiles by summing the number of checks per column and then dividing by the number of observation intervals (see

TABLE III. Number of course sections and observations for each pedagogy included in this study.

Pedagogy	Number of sections	Number of observations
Tutorials	1 Lecture	1
	19 Tutorials	1
ISLE whole class	1 Lab	1
	4 Recitations	1
	1 Lecture	2
ISLE lab only	5 Labs	1
Modeling Instruction	3	2
Peer Instruction	1	1
Context-rich problems	2 Discussions	1
	4 Labs	1
	2 Lectures	2
SCALE-UP	1	3

TABLE IV. Variance and covariance of TidyLPA model. Models 4 and 5 were not included, as they are run using the proprietary software, Mplus [26].

	Covariance		
Variance	Varying	Equal	Zero
Equal	MPLUS	3	1
Varying	6	MPLUS	2

Sec. II B for a more thorough explanation of profile creation).

For this analysis, individual class period observations are left as stand-alone COPUS profiles. We originally wanted to cluster observations into a single profile to represent a week's worth of class time for each pedagogy. This means, for example, an ISLE course that includes a lecture, recitation, and lab, would be reported as one COPUS profile that encapsulates the entire pedagogical experience for that week. Unfortunately, grouping observations in this way ensured we did not have enough data points for the LPA models to converge. As such, we left the COPUS profiles as individual class period observations, separated into various course components. For future work, we would recommend grouping observation components to represent, effectively, a "week in the life" of a student enrolled in the course. One area of contention, however, is how to do this for large enrollment courses with one lecture and multiple recitations in a way that does not overly weight one component of the course in the analysis.

A. TidyLPA in R

There are six different LPA Gaussian mixture models that can be used in the TidyLPA package in R [25], two of which are run using wrappers to the proprietary software, Mplus [26]. As such, those two models are not included in this analysis.

Each model specifies how the variance and covariance of the Gaussian mixture models is altered between iterations. For equal variance, the variance of the fit Gaussians for each profile are assumed to be equal across profiles. The covariance for inter-COPUS code relationships can be set to zero, equal between profiles, or varying between profiles. Table IV illustrates the variance and covariance combinations available, and their model number in TidyLPA.

In this paper, we show the results from model 1 analysis. Because of our relatively small dataset, models 2 and 6 fail to converge. Model 3, which assumes equal covariance between observed variables for each profile, was ultimately discarded as we felt it was not appropriate to force that condition. It would, however, be more interesting and realistic to use models 2 or 6, but we require more observations in order to get a convergent result.

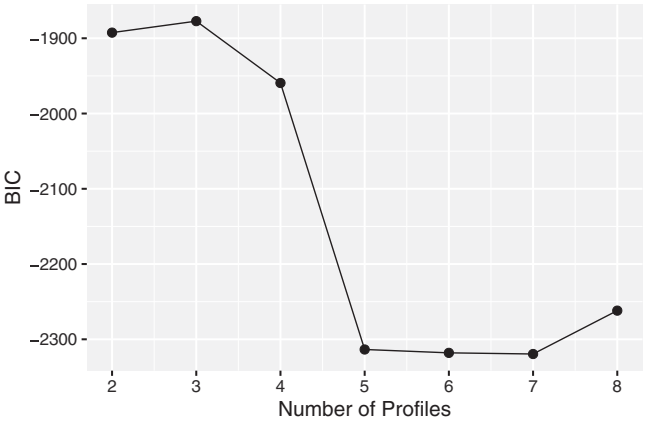

FIG. 2. BIC vs number of profiles. We see a slight increase in BIC from two (BIC -1892) to three (BIC -1877) profiles, and then a sharp drop to 4 (BIC -1959) until five (-2314) profiles. Five, six (BIC -2318), and seven (BIC -2320) profiles have nearly equal BIC values, before increasing again with eight (BIC -2262) profiles. As such, we focus on two profiles, to observe the minimum case, and five profiles in further discussion.

IV. RESULTS

We ran LPA on our data using model 1 (equal variance between profiles, zero covariance) for 2–8 profiles. This range of profiles was originally chosen because it was the same range that is used in the TidyLPA vignette [25], but was ultimately kept after investigating the results. The sorting of observations into profiles becomes nonsensical after 8 profiles.

The BIC was used as the primary indicator of model acceptability [27]. In general, the further left on the number line, the better the fit of the model. The BIC is plotted for model 1 against the number of profiles that the COPUS data was sorted into, which can be seen in Fig. 2. We present the 2 profile and 5 profile solutions here, and include the remainder in the Supplemental Material [20]. The 2 profile solution is included due to the successful sorting of observations into groups we interpreted as "interactive lecturelike" and "other," as seen in Table V. The top panel of Fig. 3 shows the box plots for each COPUS code and their associated profile groupings.

The 5 profile solution was included in this analysis because it had the best BIC with the least number of profiles. The 5 and 6 profile solutions had a BIC difference less than 5, while the 6 to 7 profile solutions had a BIC difference less than 2. As such, we interpreted models 6 and 7 to be equally good, and model 5 to be nearly as good. The BIC values are included in the caption of Fig. 2. As model 5 had the least amount of degrees of freedom and made the most sense from our interpretations of the groupings, we chose model 5 to highlight. This fit successfully sorted the COPUS observations into groups that we interpreted as interactive lecturelike, Modeling

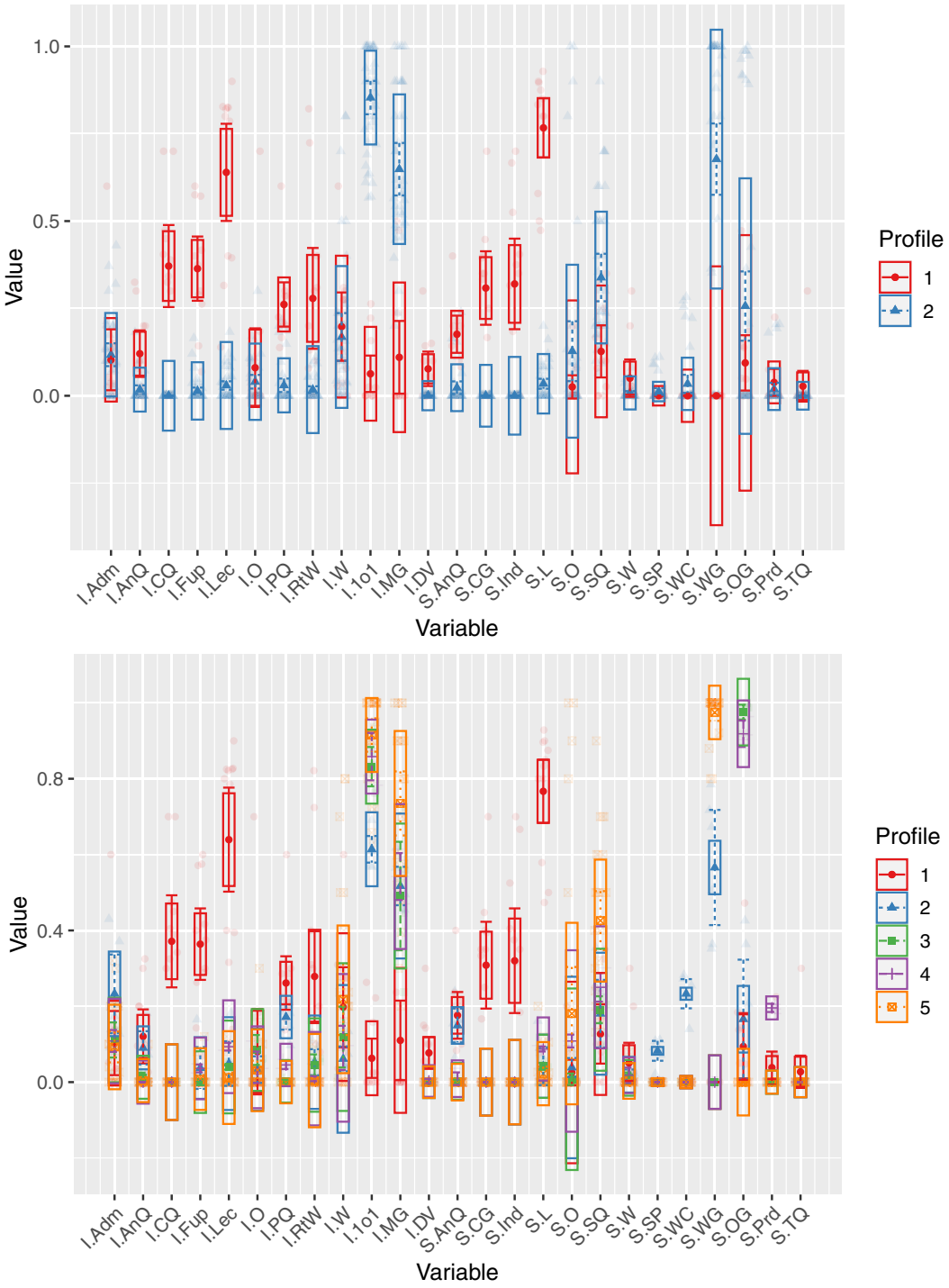

FIG. 3. Model 1 shown for 2 profiles (top panel) and 5 profiles (bottom panel). The x axis shows the full set of COPUS codes included in the analysis, while the y axis shows the fraction of class time dedicated to that code. The light shaded dots show the actual COPUS profiles, colored by their assigned profile. The boxes show the standard deviation in each assigned profile, and the bars show the confidence interval of the centroid of said profile assignment. Since we used the equal variance model, the size of the boxes is equal between classes in each COPUS code.

Instruction, ISLE labs, CRP labs, and "recitationlike" categories, as seen in Table VI. The lower panel of Fig. 3 shows the box plots for each COPUS code and their associated profile groupings.

The LPA algorithm uses hierarchical clustering to determine initial parameters, meaning the order of the data can have an effect on the resulting models. Stains *et al.* [7] combated this by shuffling their data and

TABLE V. Model 1 profile assignment, 2 profiles. Profile 1 is red on the top panel of Fig. 3, while profile 2 is blue. ISLE had two sets of observations, the whole-class implementation (WC) and the lab-only implementation (LO).

Profile 1	Profile 2	Profile 2 (cont.)	Profile 2 (cont.)	Profile 2 (cont.)
Peer instruction	ISLE LO lab 1	Modeling 1.1	Tutorials 1	Tutorials 11
SCALE-UP 1.1	ISLE LO lab 2	Modeling 1.2	Tutorials 2	Tutorials 12
SCALE-UP 1.2	ISLE LO lab 3	Modeling 2.1	Tutorials 3	Tutorials 13
SCALE-UP 1.3	ISLE LO lab 4	Modeling 2.2	Tutorials 4	Tutorials 14
ISLE WC lecture 1.1	ISLE LO lab 5	Modeling 3.1	Tutorials 5	Tutorials 15
ISLE WC lecture 1.2	ISLE WC lab 1	Modeling 3.2	Tutorials 6	Tutorials 16
CRP 1.1 lecture	ISLE WC recitation 1	CRP discussion 1	Tutorials 7	Tutorials 17
CRP 1.2 lecture	ISLE WC recitation 2	CRP discussion 2	Tutorials 8	Tutorials 18
CRP 2.1 lecture	ISLE WC recitation 3	CRP lab 1	Tutorials 9	Tutorials 19
CRP 2.2 lecture	ISLE WC recitation 4	CRP lab 2	Tutorials 10	
Tutorials lecture		CRP lab 3		
		CRP lab 4		

running the analysis 10 000 times. As this paper is proof of concept with a limited dataset, we randomized the data five times to determine if further randomization was needed. After 5 additional analyses, there were zero discrepancies between profile assignments. As such, we did not run further randomization trials. Part of the LPA output is the probability of assignment to each profile, so examining these probabilities gives a measure of certainty (or ambiguity) for the classification of each set of observations. For all trials in our analysis, the probabilities of assignment to nondesignated profiles were at 10^{-6} or less, indicating that each set of observations belonged very strongly to one profile and not to any of the others.

TABLE VI. Model 1 profile assignment, 5 profile. Profile 1 is red in the lower panel in Fig. 3, profile 2 is blue, profile 3 is green, profile 4 is purple, and profile 5 is orange. ISLE had two sets of observations, the whole-class implementation (WC) and the lab-only implementation (LO).

Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
Peer instruction	Modeling 1.1	ISLE WC lab 1	CRP lab 1	ISLE WC recitation 1
SCALE-UP 1.1	Modeling 1.2	ISLE LO lab 1	CRP lab 2	ISLE WC recitation 2
SCALE-UP 1.2	Modeling 2.1	ISLE LO lab 2	CRP lab 3	ISLE WC recitation 3
SCALE-UP 1.3	Modeling 2.2	ISLE LO lab 3	CRP lab 4	ISLE WC recitation 4
ISLE WC lecture 1.1	Modeling 3.1	ISLE LO lab 4		CRP discussion 1
ISLE WC lecture 1.2	Modeling 3.2	ISLE LO lab 5		CRP discussion 2
CRP 1.1 lecture				Tutorials 1
CRP 1.2 lecture				Tutorials 2
CRP 2.1 lecture				Tutorials 3
CRP 2.2 lecture				Tutorials 4
Tutorials lecture				Tutorials 5
				Tutorials 6
				Tutorials 7
				Tutorials 8
				Tutorials 9
				Tutorials 10
				Tutorials 11
				Tutorials 12
				Tutorials 13
				Tutorials 14
				Tutorials 15
				Tutorials 16
				Tutorials 17
				Tutorials 18
				Tutorials 19

V. DISCUSSION

A. Two profiles

If we look more closely at the top panel of Fig. 3, we can start to pick apart the differences in the assigned profiles. The two profiles that emerged effectively sorted the observations into what we interpreted as "large lecturelike" and "other." This becomes more apparent as we zoom into the individual codes. The largest separation between the two profiles occurs with the instructor lecturing (I.Lec), students listening (S. L), and the instructor having one on one conversations with students or small groups (I.1o1). If we think about how the observations were sorted (see Table V for the full list), this makes sense. While Peer instruction and SCALE-UP are pedagogies that incorporate active learning techniques, they are still typically large classes where the bulk of information transfer occurs. Pair those with the lecture components of the other pedagogies, and we can see how I.Lec and S. L were the starkest differences from the profile that encompasses lab and recitation components. The lecturelike profile that we see in our data more closely aligns with the hybrid category ("active lecture") that Stains *et al.* [7] report.

The code where profile 2 takes the lead, I.1o1, indicates that instructors spend more time engrossed in individual discussions with students outside of large lecturelike environments. This is to be expected when comparing small, typically small-group activity-centered environments to a large information transfer environment. The outlier to this generalization is Modeling Instruction; a large enrollment course with very little information transfer from the instructor via lecture, and large amounts of small group learning activities (which is likely why it was grouped with recitation and lab components).

On the other hand, there were a few codes that were nearly indistinguishable between the two profiles, including administrative tasks (I. Adm), students and instructor waiting (S. W. and I. W.), test or quiz (S.TQ), students making a prediction (S. Prd), and student presentations (S. SP). The only pedagogy that had student presentations was Modeling Instruction, when the students gathered for their white board meetings. Only one observation had a test or quiz, given in the lecture component of Tutorials. Student prediction is for when students are explicitly asked to make a prediction about the outcome of a demo or experiment, which was also a very rare occurrence in our set of observations. Meanwhile, administrative tasks, waiting around for students to finish an activity or the instructor to begin the next part of the lesson, and other unclassifiable activities (like eating a snack or taking a restroom break), seem to be universal to all pedagogies.

Codes related to clicker questions or whole-class questioning were more common in the first profile, which is to be expected, as large enrollment courses are typically the environment for this instructional tool. These codes included I.CQ (instructor administers clicker question), I.Fup (instructor follows up a question with a longer discussion), I.PQ (instructor poses a nonrhetorical question to the whole class), S.AnQ (student answers a question with the whole class listening), S.CG (students work as a group to answer a clicker question), and S.Ind (students think independently to answer a question posed to the whole class).

While the first profile had very little group worksheet activity (S.WG) outside of SCALE-UP, the second profile had a large variance due to the wide array of group activities that are encompassed in these observations. Lab sections had mostly S.OG (other group activity), Tutorials had mostly S.WG, and Modeling Instruction had a mixture of both. As such, this code forced a large variance in the first profile as well, since we used model 1 during our LPA.

B. Five profiles

Now let us look more closely at the bottom panel of Fig. 3. Profile 1 (red boxes), which included lecture components and SCALE-UP observations, had a higher usage of clicker questions (I.CQ, S.CG, and S.Ind) and subsequent follow up (I.Fup). Profile 1 also had significantly higher usage of lecture (I.Lec) and thus students listening (S. L). Also noteworthy of profile 1 is a distinct lack of one-on-one discussions (I.1o1) and moving about the room to guide discussion (I.MG).

Profile 2 (blue boxes), which holds all of the Modeling Instruction observations, is the only pedagogy that used student presentations (S.SP) and whole class discussions (S.WC). Profile 2 also shows slightly less time engaged in one-on-one discussions than the other active profiles. Less time was also spent engaged in group worksheets (S.WG) than the other active profiles.

Profiles 3 and 4, which hold the ISLE labs and the context-rich problem labs, respectively, are a little more difficult to distinguish. The most notable difference between these two lab sortings is the use of student predictions (S.Prd) in the CRP labs. In these observations, students had to predict the outcome of their experiments and discuss as a group before jumping into the activity, whereas the ISLE labs were less explicitly structured in this regard. The CRP labs were actually the first group to separate from the two-profile solution due to this code (see 3 profiles in the Supplemental Material [20]), followed by the ISLE labs in the 4 profile solution. It is also worth mentioning that the ISLE labs all go together, even though most are the stand-alone implementation and one is part of the full-class implementation. That suggests that the lab-only implementation had high fidelity to the goals of the whole-class ISLE implementation.

Finally, the recitationlike observations in profile 5 stand out with their abundance of group worksheet activities (S.WG). This profile also has elevated levels of moving

around the room to guide discussion (I.MG), as well as students raising their hands to ask questions (S.SQ).

C. Further discussion

Modeling Instruction and SCALE-UP are both studio format with a mingled lecture-lab-recitation, but were not binned in the same profiles for any number of profile fits. Unlike Modeling Instruction, the SCALE-UP class we observed still used lecture as the primary means of information transfer, whereas the bulk of information transfer in a Modeling environment is through exploration by the students and the subsequent whole class discussions. This, plus the usage of clicker questions, was the biggest difference between the seemingly similar studio-based pedagogies. As mentioned in the introduction of the pedagogies, SCALE-UP is more-so descriptive of the physical environment rather than specific activities, and lends itself well to a combination of many styles of active learning. It is therefore possible that the COPUS profile of a SCALE-UP course could vary wildly between implementations, but it is also possible that a unique SCALE-UP signature could emerge that encompasses all codes due to the physical capability to cover all codes. More observations from several institutions will be necessary to investigate further.

It is also interesting to see how all of the recitation or discussion observations were grouped in both 2 and 5 profile cases. The largest distinction between these and the lab sections was the usage of S.WG and S.OG. The bulk of recitation or discussion section activities was with small group worksheets, hence the WG "working in groups on a worksheet" designation. Meanwhile, the labs were coded as "other group activity," as COPUS does not include a code for this purpose. Without this distinction, there would likely be a lot of overlap between lab and recitation profile assignment, as the base behaviors of working in small groups while the instructor mills about the room were largely similar. For future work, we recommend exploring different observation protocols that can better capture the difference between group activities, or using both COPUS and LOPUS (Lab Observation Protocol for Undergraduate STEM, which is the lab variant of COPUS) [28].

VI. CONCLUSION

We collected COPUS observations from six high-fidelity active learning pedagogies in physics. Using these observations, we performed LPA for 2–8 latent profile groupings. Two latent profiles successfully categorized the observations into lecturelike and not lecturelike. Five latent profiles sorted the observations into lecturelike, Modeling Instruction, ISLE labs, context-rich labs, and recitationlike. With our preliminary dataset, we made a first draft classification scheme using COPUS and LPA.

This was a proof of concept study, which showed great promise for using LPA to study how COPUS profiles can be used to classify physics active learning pedagogies. Gathering additional data is an important next step, since our study included 52 total cases (sections with observed COPUS code frequencies), and this is substantially below the sample size of 500 recommended as a rule of thumb for LPA [27]. It is encouraging that our model converged and had high-probability profile assignments despite this limitation. However, more data would address this concern.

Stains *et al.* [7] suggest that at least four COPUS observations are required to get a good snapshot of each class, because day to day fluctuations had some instructors categorized into more than one cluster in their study. While we were unable to achieve this with our current iteration of data collection, we suggest future studies take this suggestion into account, and make a "combined COPUS profile," consisting of several days worth of observations that include all components of a course (like lecture, lab, and recitation combined into one observation, as if a student were encountering the entire class). This method of data collection would allow us to more directly explore the difference between pedagogies, such as how a Peer Instruction experience with separate lab and recitation differs from a SCALE-UP integrated lab-lecture-recitation experience.

Though five profiles are a good fit to the six sampled curricula, it is not quite as simple as finding one profile per pedagogy. Instead, we find that LPA can group broad components of the courses as similar (e.g., all the lecture-heavy observation sets), while still identifying differences within these components (like the split lab types in profiles 3 and 4). If more COPUS observations from other institutions and pedagogies map onto these same profiles, it suggests that they characterize broad recurring structures in active learning environments. If more or different profiles emerge as more data are added, it suggests that additional dimensions are needed to fully describe this space. In either case, the codes that distinguish the profiles—such as use of student predictions in CRP labs—help to identify distinct features between active learning pedagogies, which is the larger goal of the CALEP project.

ACKNOWLEDGMENTS

We would like to thank the institutions that participated in data collection and observation arrangements. We would also like to express our heart-felt thanks to the instructors that allowed us to infiltrate their classrooms for a week. Lastly, we would like to thank the National Science Foundation (DUE 1711017 and 1712341) for their generous support of this project.

[1] D. E. Meltzer and R. K. Thornton, Resource letter ALIP–1: Active-learning instruction in physics, Am. J. Phys. **80**, 478 (2012).

[2] *Reviews in PER Volume 1: Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), Vol. 1, https://www.compadre.org/per/per_reviews/volume1.cfm.

[3] B. A. Thacker, Recent advances in classroom physics, Rep. Prog. Phys. **66**, 1833 (2003).

[4] K. Commeford, E. Brewe, and A. Traxler, Characterizing active learning environments in physics using network analysis and classroom observations, Phys. Rev. Phys. Educ. Res. **17**, 020136 (2021).

[5] M. K. Smith, F. H. M. Jones, S. L. Gilbert, and C. E. Wieman, The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university stem classroom practices, CBE Life Sci. Educ. **12**, 618 (2013).

[6] A. L. Traxler, T. Suda, E. Brewe, and K. Commeford, Network positions in active learning environments in physics, Phys. Rev. Phys. Educ. Res. **16**, 020129 (2020).

[7] M. Stains *et al.*, Anatomy of STEM teaching in North American universities, Science **359**, 1468 (2018).

[8] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics*, 1st ed. (Prentice Hall, Upper Saddle River, NJ, 2002).

[9] R. E. Scherr and A. Elby, Enabling informed adaptation of reformed instructional materials, AIP Conf. Proc. **883**, 46 (2007).

[10] E. Etkina and A. V. Heuvelen, Investigative science learning environment - A science process approach to learning physics, in *Reviews in PER: Research-Based Reform of University Physics*, Vol. 1, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), http://www.per-central.org/document/ServeFile.cfm?ID=4988.

[11] E. Etkina, Millikan award lecture: Students of physics—Listeners, observers, or collaborative participants in physics scientific practices?, Am. J. Phys. **83**, 669 (2015).

[12] PhysPort Methods and Materials: University Modeling Instruction, https://www.physport.org/methods/method.cfm?G=University_Modeling_Instruction.

[13] M. Wells, D. Hestenes, and G. Swackhamer, A modeling method for high school physics instruction, Am. J. Phys. **63**, 606 (1995).

[14] E. Brewe, Modeling theory applied: Modeling Instruction in introductory physics, Am. J. Phys. **76**, 1155 (2008).

[15] E. Mazur, *Peer Instruction: A User's Manual*, 1st ed. (Prentice Hall, Upper Saddle River, NJ, 1997).

[16] C. Crouch, J. Watkins, A. Fagen, and E. Mazur, Peer instruction: Engaging students one-on-one, all at once, in *Reviews in PER: Research-Based Reform of University Physics*, Vol. 1, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), http://www.per-central.org/document/ServeFile.cfm?ID=4990.

[17] University of Minnesota Physics Education Research and Development: Minnesota Model for Large Introductory Courses, https://groups.spa.umn.edu/physed/Research/MNModel/MMt.html.

[18] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, Am. J. Phys. **60**, 637 (1992).

[19] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. Deardorff, R. J. Allain, S. W. Bonham, M. H. Dancy, and J. S. Risley, The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project, in *Reviews in PER: Research-Based Reform of University Physics*, Vol. 1, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), http://www.per-central.org/document/ServeFile.cfm?ID=4517.

[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.18.010113 for the box plots and grouping outputs for all but the 2 and 5 profile solutions from latent profile analysis.

[21] D. Oberski, Mixture models: Latent profile and latent class analysis, in *Modern Statistical Methods for HCI*, edited by J. Robertson and M. Kaptein (Springer International Publishing, Switzerland, 2016), pp. 275–287, 10.1007/978-3-319-26633-6_12.

[22] M. Stains and J. Harshman, COPUS Analyzer (Beta), http://www.copusprofiles.org.

[23] Mixture Models, Victor Lavrenko does an incredible job illustrating how this process works on his YouTube channel http://bit.ly/EM-alg.

[24] C. Fraley and A. E. Raftery, Bayesian regularization for normal mixture estimation and model-based clustering, J. Classif. **24**, 155 (2007).

[25] J. M. Rosenberg, P. N. Beymer, D. J. Anderson, C. J. Van Lissa, and J. A. Schmidt, tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software, J. Open Source Softwaare **3**, 978 (2018).

[26] L. K. Muthén and B. O. Muthén, *Mplus User's Guide*, 8th ed. (Muthén & Muthén, Los Angeles, CA, 1998), https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf.

[27] D. Spurk, A. Hirschi, M. Wang, D. Valero, and S. Kauffeld, Latent profile analysis: A review and "how to" guide of its application within vocational behavior research, J. Vocat. Behav. **120**, 103445 (2020).

[28] J. B. Velasco, A. Knedeisen, D. Xue, T. L. Vickrey, M. Abebe, and M. Stains, Characterizing instructional practices in the laboratory: The laboratory observation protocol for undergraduate STEM, J. Chem. Educ. **93**, 1191 (2016).