Science AAAS

www.sciencemag.org/content/359/6382/1468/suppl/DC1

Supplementary Materials for

Anatomy of STEM Teaching in North American Universities

Marilyne Stains,[1]* Jordan Harshman,[2] Megan K. Barker,[3,4] Stephanie V. Chasteen,[5] Renée Cole,[6] Sue Ellen DeChenne-Peters,[7] M. Kevin Eagan Jr.,[8] Joan M. Esson,[9] Jennifer K. Knight,[10] Frank A. Laski,[11] Marc Levis-Fitzgerald,[12] Christopher J. Lee,[13] Stanley M. Lo,[14] Lisa M. McDonnell,[4,14] Timothy A. McKay,[15] Nicole Michelotti,[16] A. Musgrove,[17] Michael S. Palmer,[18] Kathryn M. Plank,[19] Tamara M. Rodela,[4,20] Erin R. Sanders,[21] Natalie G. Schimpf,[4,20] Patricia M. Schulte,[4,20] Michelle K. Smith,[22] MacKenzie Stetzer,[23] Jackie Stewart,[24] Blaire Van Valkenburgh,[25] Erin Vinson,[22] Laura K. Weir,[4,26] Paul J. Wendel,[27] Lindsay B. Wheeler,[18] Anna M. Young[28]

*Corresponding author. Email: mstains2@unl.edu

Published 30 March 2018, *Science* **359**, 1468 (2018)
DOI: 10.1126/science.aap8892

This PDF file includes:

Authors' affiliations
Methods
References
Supplementary Text
Figs. S1 to S5
Tables S1 to S5

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/content/359/6382/1468/suppl/DC1)

Data S1
R Script

Both files can be found here: doi:10.7910/DVN/48MMF9

Authors' affiliations:

[1]Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588 USA.

[2]Department of Chemistry and Biochemistry, Auburn University, Auburn, AL 36849, USA.

[3]Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

[4]Carl Wieman Science Education Initiative, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[5]Center for STEM Learning, University of Colorado, Boulder, CO 80309, USA.

[6]Department of Chemistry, University of Iowa, Iowa City, IA 52242, USA.

[7]Department of Biology, Armstrong State University, Savannah, GA 31419, USA.

[8]Graduate School of Education & Information Studies, University of California, Los Angeles, CA 90095, USA

[9]Department of Chemistry, Otterbein University, Westerville, OH 43081, USA

[10]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

[11]Departments of Life Sciences Core Education and Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095, USA

[12]Center for Educational Assessment, University of California, Los Angeles, CA 90095, USA

[13]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

[14]Section of Cell and Developmental Biology, Program in Mathematics and Science Education, University of California San Diego, La Jolla, CA 92093, USA

[15]Physics Department, University of Michigan, Ann Arbor, MI 48109, USA

[16]University of Michigan, Ann Arbor, MI 48109, USA

[17]Chemistry Department, University of Calgary, Calgary, AB T2N 1N4, Canada

[18]Center for Teaching Excellence, University of Virginia, Charlottesville, VA 22903, USA

[19]Department of Education, Director of the Center for Teaching and Learning, Otterbein University, Westerville, OH 43081, USA

[20]Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

[21]Center for Education Innovation and Learning in the Sciences, University of California, Los Angeles, CA 90095, USA

[22]School of Biology and Ecology & Maine Center for Research in STEM Education, University of Maine, Orono, ME 04469, USA

[23]Department of Physics and Astronomy & Maine Center for Research in STEM Education, University of Maine, Orono, ME 04469, USA

[24]Department of Chemistry, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[25]Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA 90095, USA

[26]Department of Biology, Saint Mary's University, Halifax, NS B3H 3C3, Canada

[27]Department of Education, Otterbein University, Westerville, OH 43081, USA

[28]Department of Biology & Earth Science, Director of the Zoo and Conservation Science Program, Otterbein University, Westerville, OH 43081, USA

Methods

COPUS codes

COPUS codes (*1*) are provided in Table S1.

Evidence of widespread use of COPUS

- Described by a lead program director at the National Science Foundation as the number one protocol mentioned in the Improving Undergraduate STEM Education (IUSE) proposals.
- Used as a measure of impact by the Transforming Education, Stimulating Teaching and Learning Excellence (TRESTLE) project, an NSF-funded project across seven institutions and multiple STEM disciplines.
- Used by the Automated Analysis of Constructed Response (AACR) project, an NSF-funded project at six different Ph.D. granting institutions, to measure the impact that Faculty Learning Communities and collaboration on instructional activities has on biology faculty teaching practices.
- Used by the Mobile Summer Institutes on Scientific Teaching to monitor change in participants' practices overtime.
- Adopted by the Office of Information Technology at the University of Colorado Boulder as one of the primary observation protocols for providing formative feedback to faculty about classroom practices.

Validity of COPUS

Multiple studies have demonstrated that data collected through COPUS validly and reliably describe instructional practices in STEM classrooms (*2-4, 5*). COPUS describes the behaviors of instructors and students in the classroom not the quality of instruction and it is not intended to be linked to external criteria (*1*).

Face validity: Face validity, the extent to which the instrument appears to appropriately measure the targeted construct, was achieved during the development of the instrument by collecting feedback from K-12 teachers, STEM faculty, and science education specialists (*1*).

Content validity: Content validity, the extent to which the instrument measures all aspects of the targeted construct, can be assessed by looking at the overlap of codes in COPUS with codes in other observational protocols that also aim at characterizing instructional practices from a behavioral perspective. We conducted such analysis and found that COPUS codes overlap with on average 71% of codes from other observation protocols (see Table S2). This high overlap indicates that codes included in COPUS comprehensively measure instructional practices and broadly align with expert science education researchers' criteria for measuring these practices. This overlap also demonstrates that results from the COPUS will be informative to projects and universities using other observation protocols.

Criterion validity:

Predictive validity: Predictive validity, the extent to which the instrument predicts expected differences between groups, was demonstrated in a study by Connell, Donovan, and Chambers (*3*). In this study, the authors tested whether the level of use of student-centered practices impacted student learning and attitudes about biology. They compared two sections of the same course, one making extensive use of student-centered practices and one making moderate use of student-centered practices. Based on the design of these two sections, it was predicted that students who spend more time doing group work in the extensive section would have higher exam scores than students in the moderate section. Five observations were conducted in each section and were analyzed using COPUS. Analysis of COPUS data confirmed that the students in the extensive use section engaged more in group work and analysis of exam scores confirmed that these students outperformed students in the moderate use section.

Concurrent validity: Concurrent validity, the extent to which the instrument results correlate with the results from another well-established instrument implemented on the same set of data, was established in a study (*4*) in which COPUS and the Reformed Teaching Observation Protocol (RTOP; *6*) were used to analyze the same set of classroom observations. Analyses demonstrated that results from COPUS aligned with results obtained from RTOP.

Construct validity:

Convergent validity: Convergent validity, the extent to which a strong correlation exists between two instruments that are intended to measure the same construct, was established in a study (*2*) in which COPUS results were demonstrated to correlate with an established survey measuring instructional practices of STEM faculty (i.e., the Teaching Practices Inventory - TPI; *7*). Another study found a high level of agreement between faculty self-reported behaviors and COPUS observations of them in the classroom (*5*).

Inter-rater reliability for COPUS analysis

Each research team who contributed data to this project trained their COPUS coders and demonstrated high levels of inter-rater reliability. Cohen's Kappa averages *(8)* per site across all 25 codes ranged from 0.70 to 0.96. This measure was used at all but one site; that site used Jaccard similarity scores *(9)* and achieved an average of 0.97.

Study Design

This study used convenience sampling. In particular, the corresponding author enquired researchers collecting and analyzing STEM classroom practices with COPUS to voluntarily share their data. This inquiry was made by sending an email through DBER listserv. Two of the 25 institutions who volunteered had collected their data locally but also from other institutions. For example, data provided from one of the Midwest institutions came from observations of STEM faculty at that institution but also from faculty who were participating in the evaluation study of a national pedagogical workshop targeting new chemistry assistant professors; these participants came from chemistry departments at doctorate-granting universities. Fourteen of the institutions submitted their data through an online tool developed by the first author; this tool enabled researchers to have their COPUS data analyzed (It has been updated since the data collected for this study by the first two authors and can be found here: http://www.copusprofiles.org; the tool can now classify an observation into one of the seven profiles identified in this study). Data entries in this tool did not always make it clear if the course was taught by the same instructor. The numbers of faculty (N=548) thus represent a minimum. The characteristics of the sample analyzed are provided in Table S3 and S4. The 25 institutions collected the data at various points of time between Spring 2012 and Fall 2016, some for one semester, some across multiple semesters. The largest contributing institution comprises 17.1% of the data.

Data Cleaning

A total of 2,135 classrooms were observed using the COPUS. Of these classrooms, 107 were not included in the analysis because they came from non-STEM disciplines. 20 observations were also eliminated due to mistakes in data entry (e.g., no student behaviors were entered). Raw data came in a variety of forms such as aggregate summaries and two-minute-by-two-minute checkboxes. All raw data was converted to percent of behavior observed for the overall class duration. The cleaned full data set (Data S1) can be found here: URL.

Latent Profile Analysis

Latent profile analysis (LPA), also sometimes called a Gaussian mixture model, was conducted using the mclust package *(10)* in R v3.3.2 *(11)*. The goal of this procedure is to determine groups of responses (referred to as groups, classes, clusters, or profiles) where the members of the groups have similar responses to one another across multiple variables, but different responses from other groups based on those same variables. The details of the procedure are available in a paper by Fraley and Raftery *(12)*. Through the Bayesian Information Criterion (BIC) method proposed in this work, we determined that the "VVV" model (unconstrained) best fit the data (Fig. S1).

Based on Fig. S1, multiple solutions showed good fit between the 7-cluster and 11-cluster VVV models. We analyzed all of these solutions and found that the interpretations were not drastically different either in the profiles being represented nor in the proportion of observations that fell into each cluster. The 8-cluster solution only differed from the 7-cluster solution by containing a small, nonsensical cluster; the 9-cluster solution only differed from the 8-cluster solution by containing two small, nonsensical clusters; and so on. Therefore, we chose the 7-cluster solution to present because it contained the least number of small, non-interpretable clusters while maintaining a high level of empirical support.

One important consideration is that the mclust procedure involves hierarchical clustering to serve as initial parameters, which makes the procedure dependent on the order of the data. A solution was proposed *(13, 14)*, but given the large number of observations in our data, this solution still failed to produce a stable response with no means of determining which solution is (empirically) best. Similar to what should be done in hierarchical cluster analysis *(15)*, we wrote a script that ran the analysis 10,000 times, each time reshuffling the data in a random order. From each solution, we gathered the log likelihood (for evidence of a global maximum) and the average uncertainty in the assignment of participants to clusters (for evidence of a global minimum). These results are shown in Fig. S2. From these results, evidence for a global maximum log likelihood and global minimum mean uncertainty can be found (the solution represented by the point furthest down and to the right). Therefore, the solution presented in the paper is that of the solution with the greatest log likelihood and the lowest average uncertainty in cluster classifications. We also examined the 10 best solutions to see if they had significantly different interpretations. Each solution portrayed clusters very similar to those described in our paper and we have reported the sizes of these clusters in Table S5 (table shows percent of total sample classified in that cluster). This result serves as evidence that even though there are many unique solutions, many of the strongest empirical solutions point to the same interpretation: there are seven key instructional profiles that we have described in the paper. Where the solutions slightly differ is in how many observations were classified into each cluster. Because of this slight variation depending on which empirical solution is examined, we have reported a range as opposed to one value.

References

1. M. K. Smith, F. H. M. Jones, S. L. Gilbert, 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).
2. M. K. Smith, E. L. Vinson, J. A. Smith, J. D. Lewin, M. R. Stetzer, A campus-wide study of STEM courses: new perspectives on teaching practices and perceptions. *CBE Life Sci. Educ.* **13**, 624 (2014).
3. G. L. Connell, D. A. Donovan, T. G. Chambers, Increasing the Use of Student-Centered Pedagogies from Moderate to High Improves Student Learning and Attitudes about Biology. *CBE Life Sci Educ* **15**, (March 20, 2016, 2016).
4. T. J. Lund *et al.*, The best of both worlds: building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE Life Sci. Educ.* **14**, ar18 (2015).
5. J. M. Esson, K. Plank, P. Wendel, A. Young, in *New Perspectives in Science Education*. (libreriauniversitaria. it Edizioni., 2016).
6. M. Piburn, D. Sawada, Reformed Teaching Observation Protocol (RTOP) Reference Manual. Technical Report. (2000).
7. C. Wieman, S. Gilbert, The Teaching Practices Inventory: a new tool for characterizing college and university teaching in mathematics and science. *Cbe-Life Sci Educ* **13**, 552 (2014).
8. J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data. *biometrics*, 159 (1977).
9. P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat* **37**, 547 (1901).
10. C. Fraley, A. E. Raftery, T. B. Murphy, L. Scrucca, “Technical Report No. 597: mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation” (2012).
11. R Core Team, “R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing ” (R Foundation for Statistical Computing, 2016).

12. C. Fraley, A. E. Raftery, Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification* **24**, 155 (2007).
13. L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R Journal* **8**, 289 (2016).
14. L. Scrucca, A. E. Raftery, Improved initialisation of model-based clustering using Gaussian hierarchical partitions. *Advances in data analysis and classification* **9**, 447 (2015).
15. IBM. IBM Knowledge Center - Hierarchical Cluster Analysis. Available at: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/idh_clus.htm. (Accessed: 13th April 2017)

Supplementary Text

Distributions for all student and instructor behaviors

Fig. S3 shows 25 panels, one for each of the student/instructor behaviors during a class. Each behavior has been arranged in increasing order independent of all other behaviors and grey lines represent quartiles. As an example, the top left behavior (L, listening) was observed in students for 100% of the class intervals for ~40% of the sample classes. Student behaviors are in blue while instructor behaviors are in red.

The seven cluster solution

Fig. S4 describes the broad instructional styles and their associated instructional profiles. Fig. S5 shows the 7-cluster solution with all 25 COPUS behaviors for each cluster. The behaviors with colors correspond to the ones used in the cluster analysis.

Acknowledgment of COPUS coders

The authors of this manuscript would like to acknowledge all the individuals who coded the data with COPUS:

California State University Fullerton: Su L. Swarat

Otterbein University: Wendy Johnston, Matthew Grote, Jeffrey Lehman, Jeffrey Smith

University of British Columbia: Garrett Huwyler, Erica Jeffery

University of California, Los Angeles: Casey Shapiro, Michael Soh, Brit Toven-Lindsey

University of California, San Diego: Tiffany B. Luu, Kiara Miramon, Justin Tran, Juan Vargas

University of Colorado, Boulder: Rachael R. Baiduc, Jacob Brauchler, Stephanie Chasteen, Austin Chau, Vivian Lai, Tara Riedl, Shane Schwikert, Scott Sieke, Anna-Elise Smith, Clara Smith, Sarah Wise

University of Iowa: Jaime Emberger, Chamathca Malwathumullage, Kyungwoon Seo

University of Maine: Kenneth Akiha, Elizabeth Baker, Toni Barboza, Tristan Bates, Michele Benoit, Stacy Boyle, Virginia Brackett, Elizabeth Connors, Tracy Deschaine, Lauren Driscoll, Andrew Ford, Bill Freudenberger, Kathryn Priest Glidden, Marshall Haas, Kate Hayes, Beth Haynes, Teri Jergenson, Danielle Johnson, Jeremy Smith, Bob Kumpa, Lori LaCombe-Burby, Justin Lewin, Melissa Lewis, John Mannette, Nichole Martin, Lori Matthews, Nicole Novak, Jennifer Page, Patti Pelletier, Meredith Shelton, Beth Smyth-Handley, Damian Sorenson, Joanna Stevens, Rhonda Stevens, Nancy Stevick, Thomas White, and Michael Witick

University of Michigan: Hope Tambala, Robert Schoenfeld, Alexandra Soos, and Alex Lee

University of Nebraska-Lincoln: Devasmita Chakraverty, Travis J Lund, Trina Nelson, Matthew Pilarz, Kaitlyn Rosploch, Molly Undersander, Jonathan B Velasco, Trisha Vickrey

University of Northern Colorado: Jennifer Le, Rachel Lenh, Biscah Munyaka, and Stefani Ronquillo

University of Virginia: Courtney Goodloe, Enrique Aguilar, Megan Oliva, Michael MacKnight, Jonathan Ledesma, Abby Detharge, Shannon Wang, Karen Connors, Mary Shoup, Myra Reese, Deanna Lewis, Max Ober, Julie Nguyen, Katie Grant, Kpakpando Anyanwu, Emily Dodd, Sydney Ward, Katrina Eichner, Kimberly Rieder, Elizabeth Li, Echo Li, Jamie Wallenstein, Jessica Sookram, Molly Butters, Nick Shabanowitz

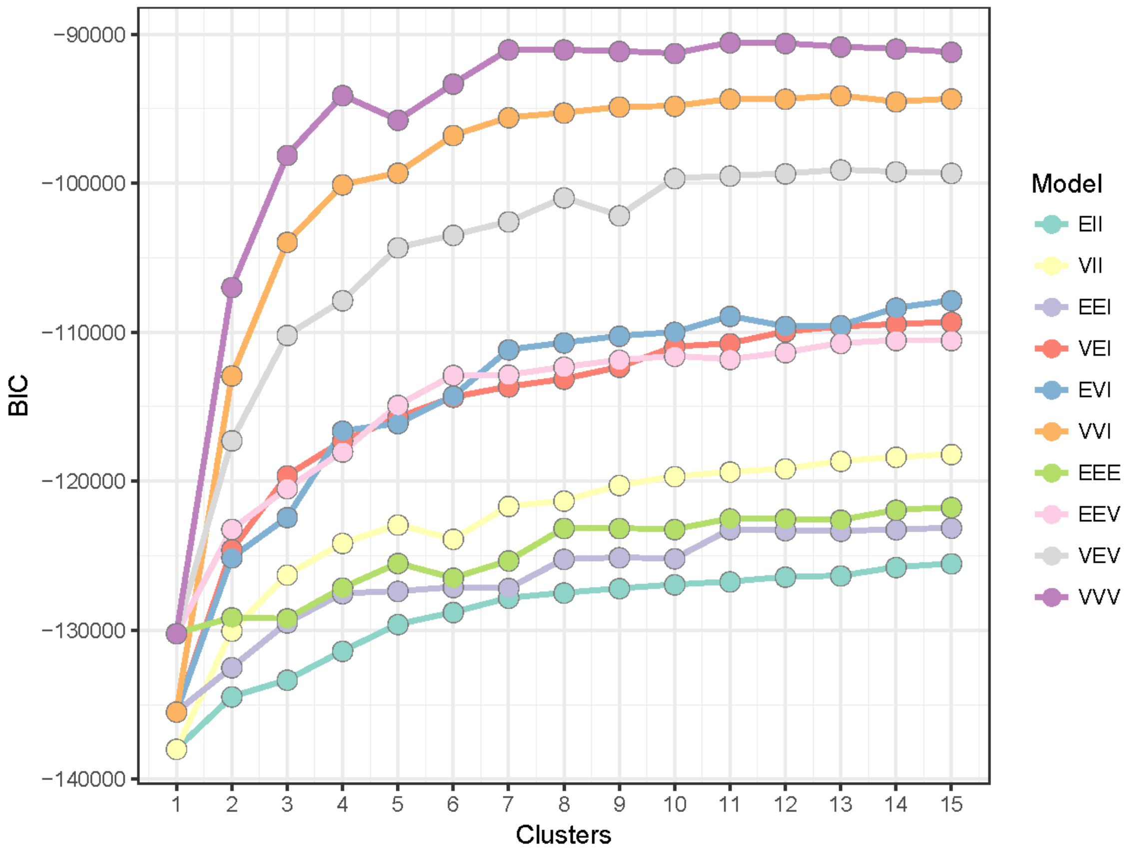

Fig. S1
Bayesian Information Criterion (BIC) values of various models to COPUS data

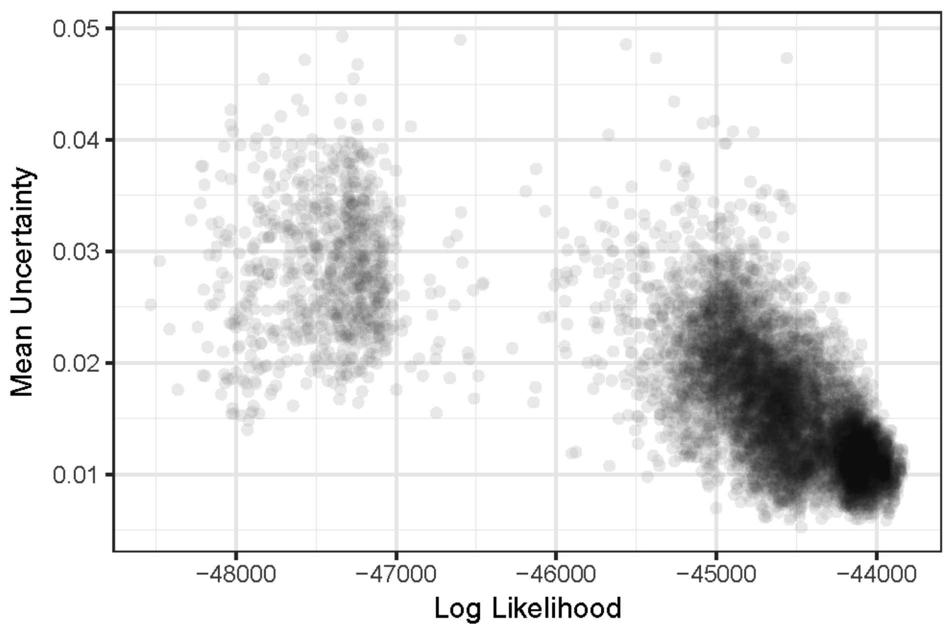

Fig. S2

Evidence for stability of Latent Profile Analysis solutions

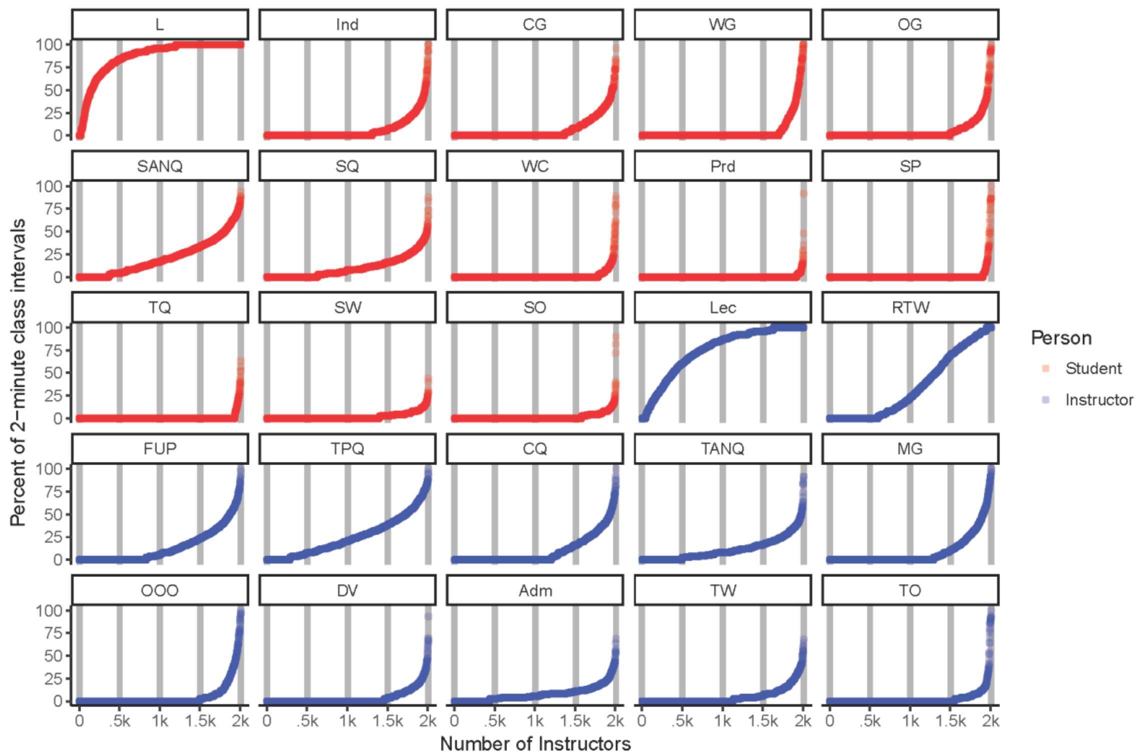

Fig. S3

Frequency of each student (blue) and instructor (red) behaviors

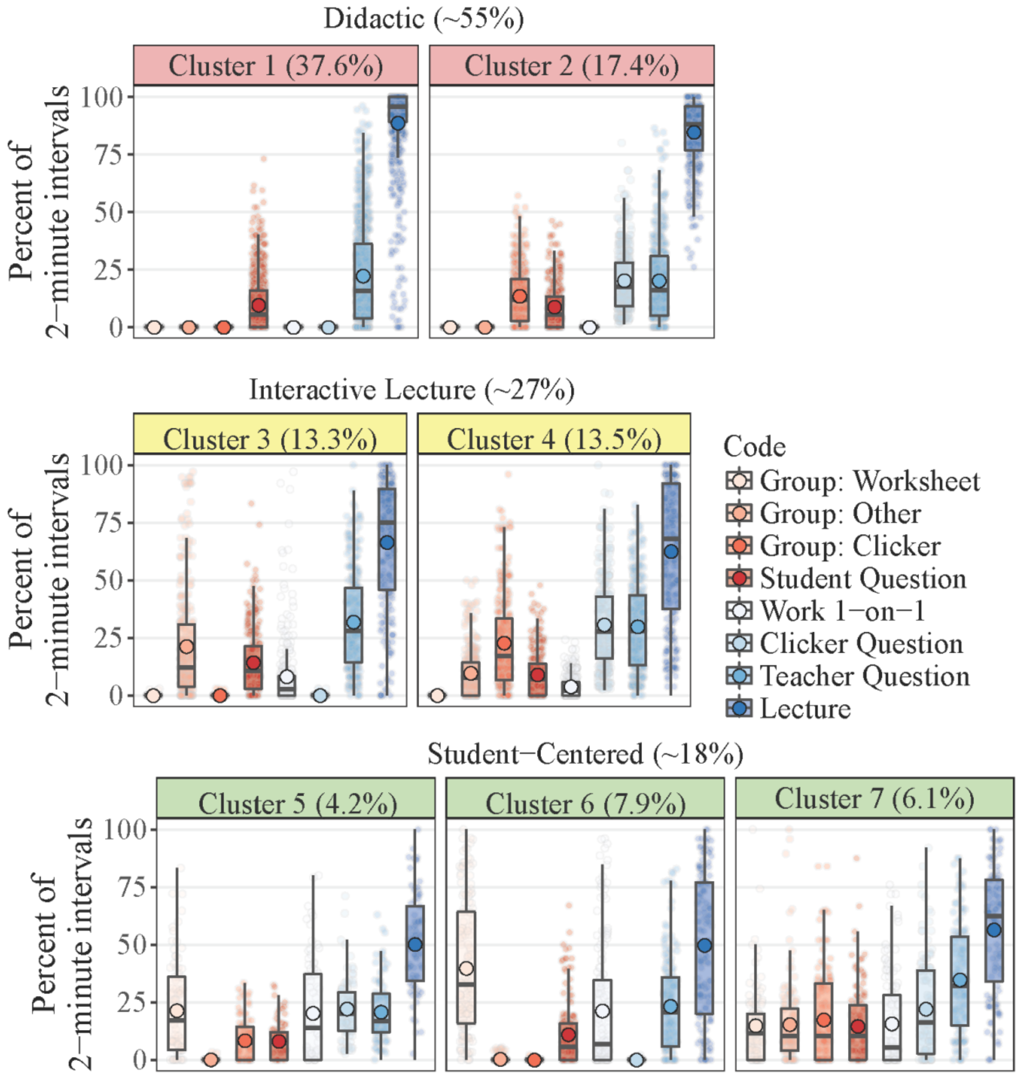

Fig. S4

Broad instructional styles and their associated instructional profiles. Each panel shows a single cluster (profile) along with the percent of observations that were classified in that cluster. Each panel shows the average (solid circle), boxplot (hollow, grey outline), and individual data points (faint points) for each of the student (shades of red) and instructor (shades of blue) behaviors.

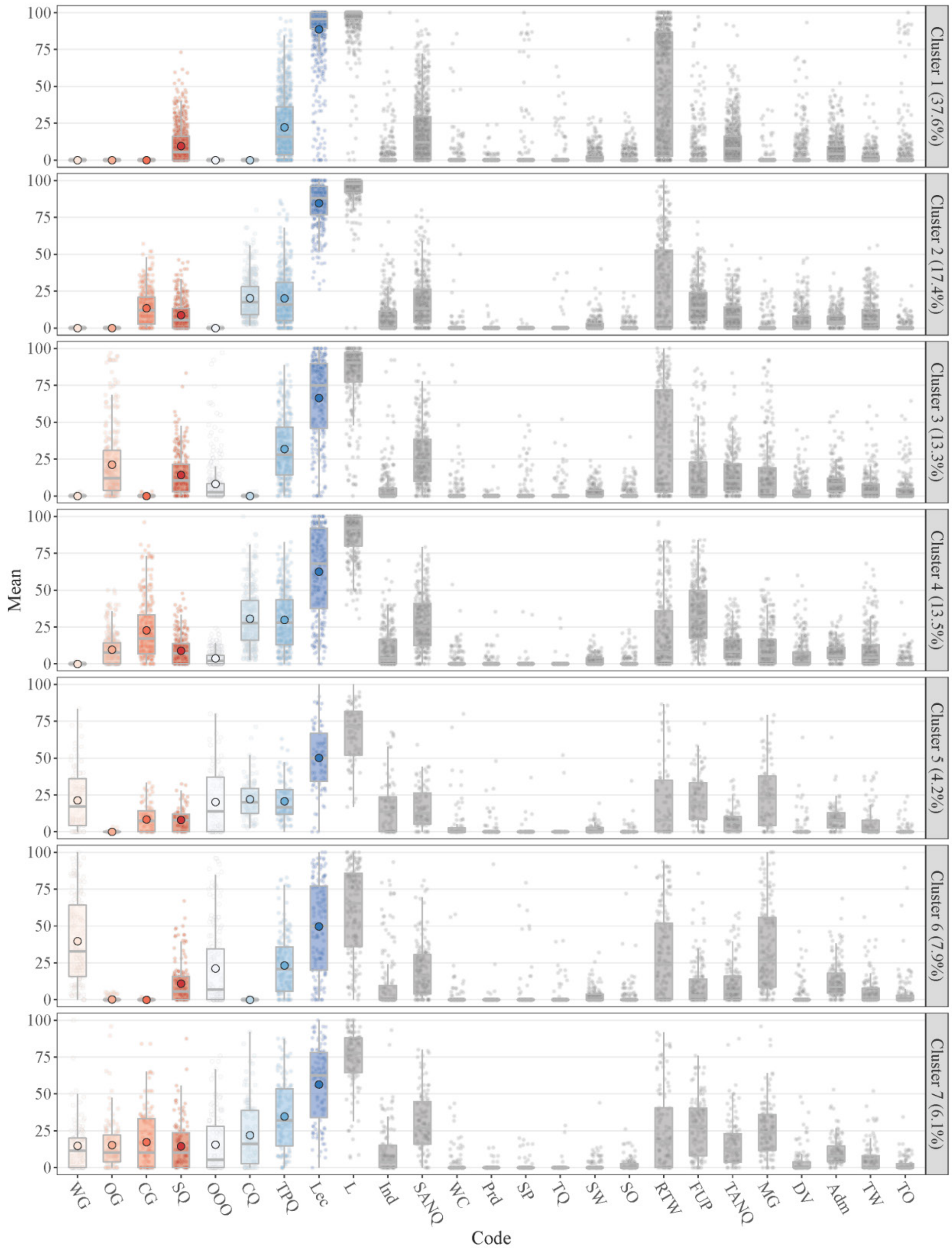

Fig. S5

Frequency of each student (shades of red) and instructor (shades of blue) behaviors that were included in the Latent Profile Analysis as well as frequency of all other student and instructor behaviors (grey) not included in the LPA within each of the seven clusters

Table S1.
COPUS codes

	Code	Definition
Students are Doing	L	Listening to instructor/taking notes, etc.
	Ind	Individual thinking/problem solving
	CG	Discuss clicker question in groups of 2 or more students
	WG	Working in groups on worksheet activity
	OG	Other assigned group activity, such as responding to instructor question
	SANQ	Student answering a question posed by the instructor with rest of class listening
	SQ	Student asks question
	WC	Engaged in whole class discussion by offering explanations, opinion, judgment, etc. to whole class, often facilitated by instructor
	Prd	Making a prediction about the outcome of demo or experiment
	SP	Presentation by student(s)
	TQ	Test or quiz
	SW	Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied, etc.)
	SO	Other – explain in comments
Instructor is Doing	Lec	Lecturing (presenting content, deriving mathematical results, presenting a problem solution, etc.)
	RTW	Real-time writing on board, doc. projector, etc.
	FUP	Follow-up/feedback on clicker question or activity to entire class
	TPQ	Posing non-clicker question to students (non-rhetorical)
	CQ	Asking a clicker question (mark the entire time the instructor is using a clicker question, not just when first asked)
	TANQ	Listening to and answering student questions with entire class listening
	MG	Moving through class guiding ongoing student work during active learning task
	OOO	One-on-one extended discussion with one or a few individuals, not paying attention to the rest of the class
	DV	Showing or conducting a demo, experiment, simulation, video, or animation
	Adm	Administration (assign homework, return tests, etc.)
	TW	Waiting when there is an opportunity for an instructor to be interacting with or observing/listening to student or group activities and the instructor is not doing so
	TO	Other – explain in comments

Table S2.
Overlap of COPUS code with other validated observation protocols organized chronologically from oldest to newest

Observation Protocol		COPUS[a]	FIA[b]	STROBE[c]	UTOP[d]	TDOP[e]	RIOT[f]	3D-LOP[g]	BERI[h]	OPAL[i]
Codes	Student	L		Student Listen	Listened to professor lecture		Talking at students		Listening	Li,Ls
		Ind		Self	Individual	DW		Tasks		
		CG		Group/student	Small group, pair	SGW,PI		Tasks	Dis/Engaged student interaction	
		WG		Group/student	Small group, pair	SGW		Tasks	Dis/Engaged student interaction	WG
		OG		Group/student	Small group, pair	SGW		Tasks	Dis/Engaged student interaction	
		SANQ	Students talk-response	Student Talk		SR	Dialoguing with student	Interactions	Engaged interaction with instructor	AnQ
		SQ	Students talk-initiation	Tally of questions - Student		SQ	Dialoguing with student	Interactions	Engaged interaction with instructor	
		WC	Students talk-initiation	Student Talk	Discussion/Whole group		Dialoguing with student/Observing	Interactions	Engaged interaction with instructor	
		Prd	Students talk-response	Student Talk					Engaged interaction with instructor	
		SP	Students talk-initiation	Student Talk	Listened to student presentations	SP	Observing			SP
		TQ			Formal Assessment	A				Adt
		SW								
		SO		Other				Miscellaneous		
	Instructor	Lec	Lecturing	Instructor/Facilitate or Talk		L,LVIS, WP	Talking at students	Lecture		Lec.Lpv
		RTW				LW				
		FUP		Instructor/Facilitate or Talk			Talking at students	Lecture		Sfu,Dfu
		TPQ	Asks questions	Tally of questions - Instructor		SOC-L,IDQ,ICQ	Dialoguing with student	Question		PQv
		CQ						Clicker question		Vt
		TANQ		Instructor/Facilitate or Talk			Dialoguing with student	Interactions		

Observation Protocol	COPUS[a]	FIA[b]	STROBE[c]	UTOP[d]	TDOP[e]	RIOT[f]	3D-LOP[g]	BERI[h]	OPAL[i]
	MG		Instructor/Facilitate or Listen/Monitor			Observing			
	OOO		Instructor/Facilitate or Talk		IND				
	DV			Demonstration	LDEM				ADV
	Adm		Instructor/Facilitate or Talk		AT	Not interacting	Administration		AdC
	TW								
	TO		Other				Miscellaneous		
Total number of codes in protocol	25	10	19	12	19*	15	7	12	17
Total number of codes in protocol also measured in COPUS		4	15	8	17	11	7	4	14
Percent code overlap between COPUS and other observation protocol		40%	79%	67%	89%	73%	100%	33%	82%

*Only focused on behaviors not technology or cognitive engagement

[a] Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (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. http://doi.org/10.1187/cbe.13-08-0154

[b] Flanders, N. A. (1974). Interaction Analysis: A Technique for Quantifying Teacher Influence.

[c] O'Malley, K. J., Moran, B. J., Haidet, P., Seidel, C. L., Schneider, V., Morgan, R. O., … Richards, B. (2003). Validation Of An Observation Instrument For Measuring Student Engagement In Health Professions Settings. *Evaluation & the Health Professions*, *26*(1), 86–103. http://doi.org/10.1177/0163278702250093

[d] Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., & Marder, M. (2012). Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program. *Unpublished paper. Southern Methodist University*.

[e] Hora, M. T., Oleson, A., & Ferrare, J. J. (2013). Teaching dimensions observation protocol (TDOP) user's manual. *Tdop.Wceruw.Org*, 28. Retrieved from http://tdop.wceruw.org/Document/TDOP-Users-Guide.pdf

[f] West, E. A., Paul, C. A., Webb, D., & Potter, W. H. (2013). Variation of instructor-student interactions in an introductory interactive physics course. *Physical Review Special Topics-Physics Education Research*, *9*(1), 010109.

[g] Laverty, J. T., Underwood, S. M., Cooper, M.M., Krajcik, J.'3S., Ebert-May, D., Jardeleza, S.E., Matz, R.L., Fata-Hartley, C.L., Posey, L.A., Caballero, M.D. . (2014). *Measuring Change in Introductory Courses with Three Dimensional Learning Analytics*. Poster presented at the Physics Education Research Conference, Minneapolis, MN. https://www.compadre.org/per/perc/2014/Detail.cfm?id=5795

[h] Lane, E. S., & Harris, S. E. (2015). A New Tool for Measuring Student Behavioral Engagement in Large University Classes. *Journal of College Science Teaching*, *44*(6), 83–91. http://doi.org/10.1371/journal.pmed.0050243

[i] Patrick, H., Ryan, A. M., Anderman, L. H., Middleton, M., Linnenbrink, L., Hruda, L. Z., ... & Midgley, C. (1997). OPAL. Observing patterns of adaptive learning: A protocol for classroom observations. *Ann Arbor: University of Michigan*.

Table S3.

Characteristics of the faculty from which observation data was collected

Country		United States				Canada	Unknown[c]
Geographical area		West[a]	Midwest[b]	South	Northeast	British Columbia	
Number of Institution		3	5	1	1	1	14
Proportion of faculty	*Biology*	6%	7%	3%	7%	5%	1%
	Chemistry	0%	12%	3%	2%	0%	3%
	Computer Science	0%	0%	3%	1%	0%	1%
	Engineering	0%	1%	7%	5%	0%	1%
	Geology	1%	1%	1%	3%	0%	7%
	Mathematics	0%	3%	4%	2%	0%	0%
	Physics	0%	5%	2%	1%	0%	2%
	Missing data	0%	1%	0%	0%	0%	0%

[a] One institution from the West collected data mostly at their own institution but also collected 11% of their observations from one biology faculty from an institution in the Northeast and 15% of their observations from one faculty in another institution in the West.

[b] One institution from the Midwest, which provided 18% of the 2008 observations, collected 40% of their observations from chemistry faculty at other institutions. These observations comprised 25% of the total number of faculty included in the data set from that institution. These faculty members represent 35 different institutions (8 in the West, 9 in the Midwest, 13 in the South, and 5 in the Northeast) with an average of 1.3±0.5 faculty per institution.

[c] Data came from an online tool that enables researchers to have their COPUS data analyzed - http://www.copusprofiles.org. Data entries did not always make it clear if the course was taught by the same instructor. The numbers of faculty thus represent minima.

Table S4.
Characteristics of classroom observations

Demographic		Classroom observations	
		Frequency	**Percent**
Discipline	Biology	591	29.4
	Chemistry	709	35.3
	Computer Science	61	3.0
	Engineering	159	7.9
	Geology	121	6.0
	Mathematics	205	10.2
	Physics	148	7.4
	Missing data	14	0.7
Course level	100 level	1,140	56.8
	200 level	294	14.6
	300 level	296	14.7
	400 level	102	5.1
	Graduate	95	4.7
	Cross-listed	7	0.3
	Missing data	74	3.7
Course Size	Small (0-50)	570	28.4
	Medium (51-100)	302	15.0
	Large (>101)	881	43.9
	Missing data	255	12.7
Classroom Layout	Fixed	757	37.7
	Flexible	380	18.8
	Missing data	871	43.4

Table S5.
Sizes of top 10 cluster solutions

Cluster	[1]*	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	Min	Max	Avg
1	37.6	37.6	37.6	37.6	37.6	37.6	37.6	37.6	37.6	37.6	37.6	37.6	37.6
2	17.4	17.5	17.5	17.5	17.5	17.5	17.5	17.4	17.4	17.5	17.4	17.5	17.5
3	13.3	13.5	13.2	13.4	13.5	13.3	13.4	13.5	13.5	12.8	12.8	13.5	13.3
4	13.5	13.5	13.7	13.6	13.5	13.3	13.2	13.6	13.5	13.6	13.2	13.7	13.5
5	4.2	4.3	4.4	4.4	4.1	4.4	5.0	4.1	4.1	4.6	4.1	5.0	4.4
6	7.9	7.5	7.8	7.8	8.0	7.8	7.8	7.3	7.4	7.8	7.3	8.0	7.7
7	6.1	6.1	6.0	5.8	5.9	6.1	5.5	6.5	6.5	6.1	5.5	6.5	6.1

* empirically “best” solution; this is the solution presented in the manuscript