Effects of the learning assistant model on teacher practice

Kara E. Gray,^{1,*} David C. Webb,² and Valerie K. Otero²

¹*Department of Physics, Seattle Pacific University, 3307 Third Avenue West,
Suite 307, Seattle, Washington 98119-1957, USA*

²*School of Education, University of Colorado Boulder, 249 UCB, Boulder, Colorado 80309-0249, USA*
(Received 21 August 2015; revised manuscript received 11 July 2016; published 1 September 2016)

Through the transformation of undergraduate STEM courses, the Colorado Learning Assistant Program recruits and prepares talented STEM majors for careers in teaching by providing them with early, sustained teaching experiences. The research reported here compares teaching practices of K-12 teachers who served as learning assistants (LAs) as undergraduates to colleagues that were certified through the same teacher certification program but did not serve as LAs. Observations of teacher practices revealed that former LAs used significantly more reformed teaching practices than their colleagues, especially in their first year of teaching. These results suggest the LA Program serves as a valuable supplement to traditional teacher certification programs.

DOI: [10.1103/PhysRevPhysEducRes.12.020126](https://doi.org/10.1103/PhysRevPhysEducRes.12.020126)

I. INTRODUCTION AND BACKGROUND

In the United States, significant concerns over relative declines in math and science majors and related careers have reenergized policy initiatives to improve the state of mathematics and science education. Over the past 25 years, the intense rhetoric that is often directed toward the quality of public schools has led to explicit or implied critiques of the quality of teachers and teaching, with mathematics and science education consistently receiving the greatest attention [1–6]. Meanwhile, research shows that among the factors within policy control, the most significant predictor of student success is the proportion of teachers in a school who have full teacher certification and a major in the subject being taught [7]. Currently, two out of three high school physics teachers in the U.S. have neither a major nor a minor in the subject [8]. Most research universities produce very few physics teachers and tend to send an implicit (and often explicit) message that secondary science teaching is not a worthy career for talented physics majors [9]. Physics departments can no longer consider physics teacher preparation the sole responsibility of schools of education, rather physics departments must play a key role in the process [10].

Research has shown that teaching strategies that respond to students' ideas during instruction improve student learning [11,12]. This has been shown to be true at both the K-12 [13–19] and the college level [20–26]. Despite the effectiveness of student-centered instruction, research has

found this to be a very difficult strategy to teach in-service teachers [13,14] and preservice teachers [27–31]. University-based programs remain the predominant teacher preparation programs in this country, and recent research has demonstrated the influence of a teacher's preparation program on teacher retention and student achievement [16,32,33]. Research is needed that articulates the features of effective STEM teacher education programs and the relationships, if any, between specific features of these programs and the conceptions and practices of the science, technology, engineering, and math (STEM) teachers who graduate from these programs. Such practices might include the use of student-centered instruction and other research-based instructional strategies. In their summary report on teacher education research, Wilson, Floden, and Ferrini-Mundy [34] offered the following recommendation:

We need more studies that relate specific parts of teachers' preparation (subject matter, pedagogy, clinical experiences) to the effects on their teaching practice, and perhaps on student achievement. Studies that compare the relative importance of specific parts of teacher preparation could be useful to those designing and revising teacher education programs (p. iv).

One widely implemented program that works to recruit and prepare possible future STEM teachers to use reformed teaching practices is the Colorado Learning Assistant (LA) Program. Among other goals, the Colorado LA Program serves as an experiential learning opportunity for undergraduates, which serves as a supplement to a university teacher certification program. Participants are given the opportunity to learn about and practice reformed teaching practices and the theories that undergird these practices. Because the participants work in introductory college courses, this program allows physics departments to play

*grayk5@spu.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

a direct role in the preparation of future teachers. The purpose of this study was to investigate whether the LA Program prepared participants who became teachers to use more effective teaching strategies. More specifically, this study attempted to answer the following research question: Are secondary math and science teachers who participated in the Colorado Learning Assistant Program more likely to use reform-based teaching practices in their high school or middle school math or science classroom than their colleagues who did not participate in this program as undergraduates?

A. Colorado learning assistant program

The Colorado Learning Assistant (LA) Program began at the University of Colorado, Boulder (CU Boulder) over a decade ago [35]. The program is focused on recruiting and preparing talented math and science teachers and improving the quality of math and science education for all undergraduates. The program has been implemented in thirteen science, math, and engineering departments at CU Boulder and is currently being emulated by over 60 two- and four-year institutions [36–41]. Each year the program at CU Boulder hires 360 LAs. LAs are undergraduate students who, through the guidance of weekly preparation sessions and a pedagogy course, facilitate discussions among groups of students in a variety of classroom settings that encourage active engagement. LAs support students by facilitating discussions about conceptual problems within the discipline. LAs focus mainly on eliciting student thinking and helping each student participate in developing a shared understanding. While most LAs are not considering careers in teaching when they first apply to the LA Program, many express an interest in teaching following their experience. This recruitment has led to a dramatic increase in the number of secondary math and science teachers certified at CU Boulder and across the state [42]. The LA Program not only recruits undergraduates to help in science and math courses but also prepares them to facilitate student learning in these transformed courses. First, LAs further their content understanding in weekly meetings with the lead instructor of the course in which they work. Second, LAs develop their pedagogical knowledge through a weekly science and mathematics education seminar. This class is attended by LAs from all departments and addresses practical techniques as well as readings from cognitive science, learning theory, and physics education research. Finally, LAs engage in actual practice as they facilitate in-class, small group learning opportunities. Research has shown that learning gains in LA supported courses are twice as high as learning gains in traditional courses using similar research-validated inventories [21,43]. Research has also shown that upper division students who took an LA supported course outperform their peers who did not take an LA supported course as freshmen [24]. Ten years

after the start of the Colorado LA Program it has been funded entirely by institutional funding sources, not by outside grants, suggesting that this program is not only successful, but also sustainable.

The main differences between the LA model and standard models for undergraduate teaching assistants include the following: (i) explicit focus on teacher recruitment and preparation, (ii) concurrent enrollment of LAs in a seminar targeted at helping them integrate content (that they are both learning as undergraduates and teaching in their LA placement), pedagogy, and their practice as an LA, (iii) a collaborative educational research program designed to evaluate the effects of the LA model, and (iv) involvement of STEM research faculty in the recruitment and preparation of future teachers.

II. METHODOLOGY

In order to investigate the teaching practices of former participants in the LA Program, a sample of teachers were observed using a structured observation protocol.

A. Participants

The sample of teachers included two groups. The first group of teachers, referred to as former LAs, consisted of certified teachers who participated in the LA Program during their undergraduate career. The second group of teachers, referred to as NonLAs, attended the same teacher certification program¹ but did not participate in the LA Program. Because of the small numbers of physics teachers, we did not have a sample large enough for our study. Furthermore, science teacher certification in Colorado is a general science certification such that teachers certified in “science” are certified to teach within any scientific discipline, which of course includes physics. We therefore chose to recruit samples of STEM teachers with the expectation that we will later be able to extend these findings to physics teachers more specifically. Teachers from these two groups were matched using one-to-one matching based on their content area (math or science), grade level (middle school or high school), and years of teaching experience. We controlled for years of teaching experience because research suggests that it is positively correlated to teacher effectiveness [34,44]. Once the groups were created, they were also found to be equivalent based on school location (urban, rural, or suburban), school context (student demographic information), number of

¹The teacher certification program these teachers attended was similar to most traditional programs. The program included education courses in the junior and senior year that culminated in a teaching internship that allowed the future teachers to practice applying what they learned. The certification program at CU Boulder has now been transformed and includes an early field experience. All of the teachers in this study graduated before the transformation process.

TABLE I. Count of teachers participating in the study.

Status	Math	Science	Total
LA	7	8	15
NonLA	7	7	14
All	14	15	29

teachers with or pursuing master's degrees in teaching, and college GPAs. The final GPAs for both groups were statistically equivalent (p value = 0.21, LA GPA = 3.38, NonLA GPA = 3.56). All teachers in the study taught within their discipline (e.g., the math teachers earned bachelor's degrees and teaching certifications in math), with the exception of one NonLA math teacher whose bachelor's degree was in science and who participated in the study for a year. Table I lists the number of teachers who participated in the project throughout its five-year span. Some teachers participated for only a year, while others participated for up to four years. Most teachers participated for several years.

B. Reformed teaching observation protocol (RTOP)

1. RTOP structure and categories

Field notes for each observation were completed using the Reformed Teacher Observation Protocol (RTOP). The RTOP is a structured observation protocol designed to quantify the extent to which a teacher's teaching practices in a single lesson align with standards-based teaching practices [45]. At the time, there were only a few teaching observation protocols available. We chose the RTOP because it was the most widely used protocol at the time of the study that is specific to math and science instruction, it has been extensively studied and evaluated, and it offers online and in-person training for observers. While the RTOP is an effective research tool, there are some limitations. The RTOP may not capture the nuances of a teachers' implementation of reformed teaching practices. This is because scores are based on an entire lesson, so individual teaching moves or instructional decisions are not captured. Therefore, a researcher can expect that two lessons with different scores would appear noticeably different to an observer, but two lessons with the same score may not have appeared exactly the same to an observer. Also, the RTOP scores a single lesson, so it does not capture the entire arc of instruction that a unit may follow. This means that if a day's lesson builds on the previous day's activities, the lesson may be scored lower than expected because a more student-centered activity from the previous day is not included in the score. We decided that any effects of these limitations would not benefit one group of teachers over the other and would only underestimate differences in reformed teaching practices between the groups.

The RTOP protocol is made up of 25 statements that are scored from 0 to 4, where zero indicates that the practice described in the statement was not existent in the class period and four indicates that the statement was characteristic of the class period [46]. The total score for an observation is found by summing the scores from the 25 statements. The 25 RTOP statements can be divided into three categories and four subcategories. These categories are lesson design and implementation, content, and classroom culture. The lesson design and implementation category looks at whether a lesson was designed to value students' ideas. The content category is divided into two subcategories: content propositional knowledge focuses on the teachers' understanding of the content and the level of abstraction and connections to the real world used within the lesson, while content procedural knowledge focuses on the processes students are asked to use such as predictions, representations, and the evaluation of claims. This second subcategory is similar to scientific reasoning or mathematical thinking. The classroom culture category is also divided into two subcategories: communicative interactions and student-teacher relationships. Communicative Interactions focuses on the extent to which students spend a lot of time communicating with each other and the degree to which these discussions are critical to the generation of and reflection on the key topics of the lesson. The student-teacher relationships category focuses on how teachers value student ideas and support students in the learning process. For each category and subcategory, the observer scores the five statements that operationalize that category.

2. RTOP factors

Using factor analysis during the development of the RTOP the developers also found five factors for the RTOP [47]. Some of these factors align with RTOP categories while others are unique. These factors are inquiry, propositional content knowledge, pedagogical content knowledge, community of learners, and reformed teaching. The inquiry factor considers how aligned the lesson was with the practices of inquiry teaching, such as having students investigate on their own; develop predictions, explanations, and representations of data; and reflect on and assess their work. The propositional content knowledge factor includes three of the five statements included in the propositional content knowledge category and focuses only on the content knowledge presented in the lesson. The pedagogical content knowledge factor measures the extent to which teachers elicit and use students' prior knowledge in their lesson design and in their teaching, as well as the extent to which they encourage students to value rigor and alternative explanations. The community of learners factor includes statements that focus on the extent to which the classroom culture encourages students to work together while the teacher acts as a listener and resource. The final factor, reformed teaching, looks at whether the teacher

encourages student exploration as well as the organization and abstraction of ideas.

The categories and factors represent two different methods of dividing the RTOP items. The categories represent a theoretical or *a priori* division of the items while the factors represent an empirical division of the items. A score for each category or factor is calculated by averaging across the scores for each statement included in that factor or category. An average is used because the number of statements in each factor varies across the factors. While the factors and categories of the RTOP are aligned with the goals and values of both the Colorado Learning Assistant Program and the teacher preparation program, the RTOP is not used in either program. Therefore, none of the teachers who participated in the study were familiar with the RTOP.

3. RTOP training

Our research group met each semester to establish reliability in RTOP scoring. These meetings included watching classroom videos (not from the teachers included in the study) and discussing the scoring. The videos used to monitor and improve interrater agreement were provided by the RTOP website and selected from publically available resources (e.g., Annenburg Learner [48]). Acceptable reliability was defined as all members' scores being within five points of each other for the overall score and not differing by more than a single point on any statement. Every time a new member was added to the research team, the entire group participated in the video-based RTOP calibration until acceptable reliability was reached. In addition, the first real-time observation for new members was completed with a more experienced team member, and was followed by in-person mediation to discuss any differences in scores and to agree upon a final score for the observed lesson. Over the course of the study nine education graduate students and three university faculty performed observations. All members of the research team were knowledgeable in reformed STEM teaching practices.

C. Data collection

Each teacher was observed at least twice and usually 3 times over the course of the school year by at least two researchers from the project. Members of the research team were not aware of whether the teachers they were observing were prior participants in the LA Program, unless a teacher mentioned this in passing. To mitigate the effects of this possibility, at least two observers observed each teacher each year so a single researcher did not complete all of the observations for a teacher. References to the LA Program were only connected to the teacher data during the analysis phase. A researcher observed a single class for the entire class period. Observations were scheduled with the teachers so that they were aware of when researchers

TABLE II. Number of observations by years of teaching experience.

		Years of teaching experience			Total
		1	2	3+	
LA	Math	17	10	11	38
	Science	21	21	15	57
	Total	38	31	26	95
NonLA	Math	18	14	9	41
	Science	20	12	10	42
	Total	38	26	19	83
Grand Total		76	57	45	178

would be attending their class, though most teachers claimed to not having made any changes to their lessons for the observation. Over the five years of the study the team completed 178 observations. The number of observations are organized in Table II by the years of teaching experience (year 1 corresponds to teachers in their first full year of teaching).

D. Alternative data analysis

The analysis described above follows the RTOP analysis described in the RTOP reference manual [47], previous RTOP studies [16,49–51], and other educational studies [52–54]. While this method provides a succinct summary of the quantitative data collected during each observation, it overlooks a key feature of this type of data. These types of studies require researchers to rate teaching practices based on a multipoint scale. The scales used on these items are not interval scales, but are ordinal scales. This means that while a score of 3 is greater than a score of 2, it is not true that the difference between a 2 and a 3 is equal to the difference between a 2 and a 1 or a 1 and a 0. Therefore, averaging these scores disregards the distinct meaning of the scale and can provide an incorrect interpretation of the data. For instance, the RTOP provides a distinct meaning for each of the five possible scores used for each item [46]. According to this scale a 0 means there was no evidence of the practice described in the statement. A 1 means that the practice occurred once during the lesson but was implemented poorly. A 2 means the practice happened more than once but was implemented poorly. A 3 means the practice was implemented pretty well during the class period. A 4 means that the statement is characteristic of the class.

In order to account for this feature of the RTOP scale, we reanalyzed the RTOP data by counting the number of times each score was given during an observation. Therefore, a single observation now has five scores (the number of 0's, the number of 1's, etc.) instead of a single score. We then averaged the counts for each of the five scores across all the observations for LAs and NonLAs.

FIG. 1. Overall RTOP scores for LAs and NonLAs disaggregated by the subject observed. Error bars represent plus and minus 1 standard error on the mean. (a) $p = 0.005$ and (b) $p = 0.001$.

III. RESULTS

In this section we present overall and construct specific results from the use of the RTOP in 178 observed lessons. All of the results are based on averaging over the number of observations. The results are organized to address the research question: How do K-12 teachers, who had been former LAs, compare to K-12 teachers who had not been former LAs in terms of their teaching practices as measured by the RTOP? In addition to the comparison of group means and related statistical tests, we also include an alternative analysis that we feel is more appropriate for a comparative analysis of ordinal data.

A. Overall RTOP scores

Figure 1 displays the average RTOP scores for all of the math and science teachers observed during the course of the study (see the Tables III–VII in the Appendix for tabulated values from all figures). The results show that former LAs have higher average RTOP scores (math-LA = 59.1; sci-LA = 58.2) than their colleagues (math-NonLA = 45.2; sci-NonLA = 44.8). These results are statistically significant for both math and science ($p = 0.005$ and $p = 0.001$, respectively) (Throughout this paper we have provided tabulated p values rather than comparisons with threshold values. This aligns with recent trends in social science research intended to provide the reader with more complete statistical information. We have maintained the standard practice of identifying results with $p < 0.05$ as “statistically significant” and results with $p > 0.05$ as “not statistically significant”). Error bars represent ± 1 standard error on the mean.

B. RTOP scores by categories and factors

Figure 2 displays the RTOP scores for all teachers in the study for each of the five categories of the RTOP.

FIG. 2. Average statement response for the five categories of the RTOP. Error bars represent plus and minus 1 standard error on the mean. (a) $p = 0.02$, (b) $p = 0.01$, (c) $p = 0.005$, (d) $p = 0.001$, (e) $p = 0.001$, and (f) $p = 0.0003$.

All categories show the same trend; former LAs’ scored higher than their colleagues. These differences are statistically significant except for the differences between former LA and NonLA math teachers in the propositional knowledge category ($p = 0.18$). The differences are statistically significant for math teachers in the lesson design and implementation ($p = 0.01$), procedural knowledge ($p = 0.005$), communicative interactions ($p = 0.001$), and student-teacher relationships ($p = 0.01$) categories. The differences were statistically significant for science teachers in the lesson design and implementation ($p = 0.02$), propositional knowledge ($p = 0.001$), procedural knowledge ($p = 0.0003$), communicative interactions ($p = 0.02$), and student-teacher relationships ($p = 0.02$) categories.

The following analysis considers the same data organized according to the RTOP factors discussed previously. While this section presents a different organization of the

FIG. 3. Average statement response for the five factors of the RTOP. Error bars represent plus and minus 1 standard error on the mean. (a) $p = 0.03$, (b) $p = 0.02$, (c) $p = 0.01$, (d) $p = 0.003$, (e) $p = 0.002$, and (f) $p = 0.001$.

FIG. 4. Overall RTOP scores disaggregated by teachers' experience. Error bars represent plus and minus 1 standard error on the mean. (a) $p = 0.01$, (b) $p = 0.002$.

data, the results remain similar. Figure 3 shows the average statement response for each of the five factors of the RTOP. Scores are organized by LA and NonLA and by the subject observed. In every factor, the LAs score higher than their colleagues. These results are statistically significant for every factor except math teachers' content propositional knowledge factor ($p = 0.22$). The results are statistically significant for math teachers in the Inquiry ($p = 0.01$), content pedagogical knowledge ($p = 0.01$), community of learners ($p = 0.003$), and reformed teaching ($p = 0.02$) factors. The results are statistically significant for science teachers in the Inquiry ($p = 0.002$), content propositional knowledge ($p = 0.001$), content pedagogical knowledge ($p = 0.01$), community of learners ($p = 0.01$), and reformed teaching ($p = 0.03$) factors.

C. RTOP scores by years of teaching experience

The results presented above include all observations completed throughout the five years of the study. Figure 4 shows the average overall RTOP scores for LAs and NonLAs disaggregated by years of teaching experience. In the figure, first year refers to all teachers in their first full year of teaching. Because this study focused on teachers' practice during their induction years, and most districts and researchers consider the induction or probation years to be the first two years of an initial teaching contract, we chose to group teachers in years 3 through 5 of teaching in our analysis. The data in Fig. 4 show the same trend presented earlier of LAs scoring higher than their colleagues on the overall RTOP score, though the results are only statistically significant for teachers' first year of teaching ($p = 0.002$)

FIG. 5. Analysis of all RTOP responses for LAs and NonLAs using statement score counts. Error bars represent plus and minus 1 standard error on the mean. (a) $p = 0.02$, (b) $p = 0.002$, (c) $p = 0.0004$, (d) $p = 0.00002$.

and for three or more years of teaching ($p = 0.01$). There is no statistical difference between teachers' scores in their second year of teaching ($p = 0.127$).

D. RTOP scores using alternative analysis

After analyzing the data using the traditional data analysis methods for RTOP data, we reanalyzed the data using the alternative analysis described previously. These data are presented in Fig. 5. The maximum possible count for any of the statement scores in Fig. 5 would be 25, if an observation received the same score on all statements of the RTOP. A completely random scoring of the observation would result in an average count of 5 for each statement score in Fig. 5. Based on Fig. 5 we see that the NonLAs have more low scores (0, 1) on their observations than LAs while the LAs have statistically more high scores (3, 4) than the NonLAs. There is no statistical difference in the average number of medium scores (2) for LAs and NonLAs. This suggests that the LAs' teaching practices tend to be more aligned with research-based teaching practices as their classroom practices are more likely to be somewhat characteristic or very characteristic of the RTOP statements. The NonLAs are more likely to not enact a teaching practice described on the RTOP, or to do it only once, during a class period compared to the LAs. When NonLAs do enact a practice described on the RTOP, the observation of their practice is more likely to be marked as a low score for that statement than a high score.

Overall, this categorical analysis shows similar results to the previous analysis of means. The LAs tend to implement specific reformed teaching practices well while NonLAs are more likely to not implement specific reformed teaching practices at all or rarely as compared to LAs. This is a more specific result than the previous analysis offered which suggested that both LAs and NonLAs may be implementing reformed teaching practices equally often in their

classrooms but that the LAs' implementation may tend to be better.

IV. CONCLUSIONS AND IMPLICATIONS

While the data show that former LAs tend to use more reformed teaching practices in their secondary classrooms than their colleagues, it is not able to explain why this happens. Therefore we propose four possible explanations for the LAs' increased use of reformed teaching practices. First, it is possible that LAs have higher RTOP scores because they had more time to practice their teaching than their colleagues. On average, LAs spend 60 hours in a classroom each semester. This experience is prior to their entry in the teacher certification program and is therefore in addition to the teaching experience they receive as part of the certification program. If this is the accurate explanation, then the LA Program can play an important role in teacher preparation by providing an opportunity for future teachers to receive experience teaching in a context that does not require them to leave their college campus, is discipline specific, and supports the learning of their peers. A second explanation for LAs' higher RTOP scores is that the LA Program recruits students who are more likely to use reformed teaching practices. Since the LA Program has significantly increased the number of STEM majors earning secondary teaching certifications [42], this explanation would mean that the LA Program is able to attract reform-oriented STEM majors who had not previously considered careers in teaching. A third possible explanation for LAs' higher RTOP scores is that the experiential nature of the LA Program allows LAs to integrate their knowledge of pedagogy with their experiences in the classroom. This integration would allow for a deeper learning that LAs are able to transfer to their secondary math and science classrooms. If this explanation is accurate, then the three-pronged model of the LA Program provides a critical opportunity for future teachers to reflect on their teaching and to integrate theories of teaching and learning into their practices. A final possible explanation is that when pedagogical instruction is framed from the viewpoint of helping students learn rather than from the viewpoint of learning how to teach (as is often the case in methods courses in teacher certification programs), it is easier to understand and adopt reform-based teaching practices [55].

The Colorado Learning Assistant Program is not intended to be a stand-alone teacher preparation or certification program. Yet, as this study indicates, the addition of

this experience to teachers' early certification experience better prepared them to implement the reformed teaching practices for which education research and policy are calling. This program therefore provides a model for teacher preparation that indicates the importance of situating the teaching of pedagogical content within the STEM content and sustained teaching opportunities in classrooms using reformed teaching practices. By preparing STEM teachers to implement reformed-teaching practices, especially during their induction years of teaching secondary mathematics and science, we are working to provide high quality learning opportunities to all math and science students.

While this work addresses an important question about the LA Program, several questions remain. First, while this work addressed teachers' practices, future work will investigate the proposed explanations for why LAs differed from NonLAs in their use of reformed teaching practices. This will require studying LAs during their undergraduate experience and before they begin their K-12 teaching career. As the program has expanded throughout the nation and throughout the world, many other universities are studying the LA experience and generating explanations for why they differ from their peers.² Finally, future work will consider the effectiveness of the LA Program as it is emulated at other universities. The study of other implementations of the LA Program will help to address the above lines of inquiry as well as help to clarify the critical elements of the LA model.

ACKNOWLEDGMENTS

We gratefully acknowledge the many people who helped observe teachers including Erin Allman, Stephanie Barr, Kim Geil, Ian Her Many Horses, Steve Iona, May Lee, Deb Morrison, Michael Ross, and Bud Talbot. We are grateful for the support of the research teams at the University of Colorado-Boulder and Seattle Pacific University. This work was supported by NSF Grant No. ESI-0554616.

APPENDIX: TABULATED VALUES FROM FIGURES

The following tables present the data from Figs. 1 through 5 in tabulated form.

²For a list of papers associated with the LA model, go to laa.colorado.edu and log in for access to the resources and other sites. Resources can also be directly accessed at <https://sites.google.com/a/colorado.edu/la-resources/>

TABLE III. Overall RTOP scores for LAs and NonLAs disaggregated by the subject observed.

	LA			NonLA			<i>p</i> value
	Mean	SD	N	Mean	SD	N	
Math	58.5	19.9	38	46.3	18.7	41	0.0052
Science	56.7	18.2	57	45.7	15.3	42	0.0011

TABLE IV. Average statement response for the five categories of the RTOP.

		LA			NonLA			<i>p</i> value
		Mean	SD	N	Mean	SD	N	
Math	Lesson design and Implementation	2.0	1.0	38	1.5	0.8	41	0.013
	Propositional knowledge	2.8	0.7	38	2.6	0.7	41	0.183
	Procedural knowledge	1.9	1.0	38	1.3	0.9	41	0.005
	Communicative interactions	2.2	0.8	38	1.6	0.8	41	0.001
	Student-teacher relationships	2.8	0.9	38	2.3	0.9	41	0.010
Science	Lesson design and implementation	1.8	1.0	57	1.4	0.8	42	0.021
	Propositional knowledge	2.9	0.6	57	2.5	0.7	42	0.001
	Procedural knowledge	1.7	0.9	57	1.1	0.7	42	0.000
	Communicative interactions	2.1	0.9	57	1.6	0.8	42	0.018
	Student-teacher relationships	2.8	0.8	57	2.4	0.8	42	0.018

TABLE V. Average statement response for the five factors of the RTOP.

		LA			NonLA			<i>p</i> value
		Mean	SD	N	Mean	SD	N	
Math	Inquiry	1.9	1.0	38	1.3	0.9	41	0.011
	Content propositional knowledge	2.6	0.8	38	2.4	0.8	41	0.219
	Content pedagogical knowledge	1.9	1.0	38	1.3	0.9	41	0.008
	Community of learners	2.9	0.9	38	2.3	0.9	41	0.003
	Reformed teaching	1.9	1.0	38	1.4	0.9	41	0.021
Science	Inquiry	1.6	0.9	57	1.1	0.7	42	0.002
	Content propositional knowledge	2.9	0.7	57	2.4	0.8	42	0.001
	Content pedagogical knowledge	1.9	1.0	57	1.4	0.8	42	0.009
	Community of learners	2.8	0.9	57	2.3	0.9	42	0.012
	Reformed teaching	1.8	0.9	57	1.4	0.8	42	0.034

TABLE VI. Overall RTOP scores for LAs and nonLAs organized by years of teaching.

	LA			NonLA			<i>p</i> value
	Mean	SD	N	Mean	SD	N	
1st year	57.9	21.5	38	43.2	19.0	38	0.002
2nd year	57.6	17.4	31	51.2	14.3	26	0.127
3 + years	56.5	16.9	26	44.5	15.1	19	0.012

TABLE VII. Analysis of all RTOP responses for LAs and NonLAs using statement score counts.

	LA			NonLA			<i>p</i> value
	Mean	SD	N	Mean	SD	N	
0	3.3	3.6	95	5.1	4.0	83	0.0015
1	4.3	3.0	95	5.9	2.9	83	0.0004
2	5.2	2.9	95	5.4	2.6	83	0.6506
3	6.3	3.3	95	5.3	2.8	83	0.0240
4	5.9	4.7	95	3.4	3.4	83	0.0000

- [1] D. P. Gardner *et al.*, *A Nation At Risk: The Imperative For Educational Reform. Report of the National Commission on Excellence in Education* (Washington, DC, 1983).
- [2] National Research Council, *Everybody Counts. A Report to the Nation on the Future of Mathematics Education* (National Academies Press, Washington, DC, 1989).
- [3] U.S. Department of Education, *Before It's Too Late: A Report to the Nation from the National Commission on Mathematics and Science Teaching for the 21st Century* (Education Publications Center, Jessup, 2000).
- [4] National Science Board, *The Science and Engineering Workforce: Realizing America's Potential* (National Science Foundation, Arlington, VA, 2003).
- [5] Domestic Policy Council, *American Competitiveness Initiative: Leading the World in Innovation* (Office of Science and Technology Policy, Washington, DC, 2006).
- [6] National Research Council, *A Framework for K-12 Science Education, Practices, Crosscutting Concepts, and Core Ideas* (The National Academies Press, Washington, D.C., 2011).
- [7] L. Darling-Hammond and P. Youngs, Defining "Highly Qualified Teachers": What does "Scientifically-Based Research" actually tell us?, *Educ. Res.* **31**, 13 (2002).
- [8] M. Neuschatz and M. McFarling, *Reaching the Critical Mass: The Twenty Year Surge in High School Physics. Findings from the 2005 Nationwide Survey of High School Physics Teachers. AIP Report.* (American Physical Society, College Park, MD, 2008).
- [9] D. E. Meltzer, M. Plisch, and V. Stamatidis, *Transforming the Preparation of Physics Teachers: A Call to Action. A Report by the Task Force on Teacher Education in Physics (T-TEP)* (American Physical Society, College Park, MD, 2012).
- [10] T. Hodapp, J. Hehn, and W. Hein, Preparing high-school physics teachers, *Phys. Today* **62**, 40 (2009).
- [11] National Research Council, *Adapting to a Changing World—Challenges and Opportunities in Undergraduate Physics Education* (National Academies Press, Washington, DC, 2013).
- [12] National Research Council, *How People Learn: Brain, Mind, Experience, and School* (National Academy Press, Washington, DC, 2000).
- [13] P. Black and D. William, Assessment and classroom learning, *Educ. Assess.* **5**, 7 (1998).
- [14] V. J. Shute, Focus on formative feedback, *Rev. Educ. Res.* **78**, 153 (2008).
- [15] R. Geier, P. C. Blumenfeld, R. W. Marx, J. S. Krajcik, B. Fishman, E. Soloway, and J. Clay-Chambers, Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform, *J. Res. Sci. Teach.* **45**, 922 (2008).
- [16] C. Jong, J. J. Pedulla, E. M. Reagan, Y. Salomon-Fernandez, and M. Cochran-Smith, Exploring the link between reformed teaching practices and pupil learning in elementary school mathematics, *School Sci. Math.* **110**, 309 (2010).
- [17] S. L. Adamson, D. Banks, M. Burtch, F. Cox, E. Judson, J. B. Turley, R. Benford, and A. E. Lawson, Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement, *J. Res. Sci. Teach.* **40**, 939 (2003).
- [18] T. P. Carpenter, E. Fennema, P. L. Peterson, C. Chiang, and M. Loef, Using knowledge of children's mathematics thinking in classroom teaching: an experimental study, *Am. Educ. Res. J.* **26**, 499 (1989).
- [19] National Research Council, *America's Lab Report: Investigations in High School Science* (The National Academies Press, Washington, DC, 2006).
- [20] M. A. Ruiz-Primo, D. Briggs, H. Iverson, R. M. Talbot, and L. A. Shepard, Impact of undergraduate science course innovations on learning, *Science* **331**, 1269 (2011).
- [21] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [22] L. C. McDermott, Millikan lecture 1990: What we teach and what is learned - closing the gap, *Am. J. Phys.* **59**, 301 (1991).
- [23] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- [24] S. J. Pollock, Longitudinal study of student conceptual understanding in electricity and magnetism, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020110 (2009).
- [25] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410 (2014).

- [26] S. L. Laursen, M.-L. Hassi, M. Kogan, and T. J. Weston, Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study, *J. Res. Math. Educ.* **45**, 406 (2014).
- [27] V. K. Otero and M. J. Nathan, Preservice elementary teachers' views of their students' prior knowledge of science, *J. Res. Sci. Teach.* **45**, 497 (2008).
- [28] G. A. Buck, A. Trauth-Nare, and J. Kaftan, Making formative assessment discernable to pre-service teachers of science, *J. Res. Sci. Teach.* **47**, 402 (2010).
- [29] M. Lampert and F. Graziani, Instructional activities as a tool for teachers' and teacher educators' learning, *Elem. School J.* **109**, 491 (2009).
- [30] E. Kazemi, M. Franke, and M. Lampert, Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious instruction, in *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia*, edited by R. Hunter, B. Bicknell, and T. Burgess (MERGA, Palmerston North, NZ, 2009), pp. 11–30.
- [31] J. Thompson, M. Windschitl, and M. Braaten, Developing a theory of ambitious early-career teacher practice, *Am. Educ. Res. J.* **50**, 574 (2013).
- [32] D. J. Boyd, P. L. Grossman, H. Lankford, S. Loeb, and J. Wyckoff, Teacher preparation and student achievement, *Educ. Eval. Policy Anal.* **31**, 416 (2009).
- [33] S. M. Wilson, How can we improve teacher quality?, *Phi Delta Kappan* **93**, 64 (2011).
- [34] S. M. Wilson, R. E. Floden, and J. Ferrini-mundy, *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations* (Center for the Study of Teaching and Policy, Seattle, 2001).
- [35] V. K. Otero, Nationally Scaled Model for leveraging course transformation with physics teacher preparation, in *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices*, edited by C. Sandifer and E. Brewster (American Physical Society, College Park, MD, 2015), pp. 107–116.
- [36] E. Price, The CSUSM learning assistants program, *APS Forum Educ. Newsl.* **31** (2013).
- [37] N. McHenry, A. Martin, A. Castaldo, and D. Ziegenfuss, Learning assistants program: Faculty development for conceptual change, *Int. J. Teach. Learn. Higher Educ.* **22**, 258 (2010).
- [38] E. W. Close, L. Seeley, A. D. Robertson, L. S. DeWater, and H. G. Close, Seattle pacific university: Nurturing physics teachers at a small liberal-arts school, in *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices*, edited by E. Brewster and C. Sandifer (American Physical Society, College Park, MD, 2015), pp. 37–49.
- [39] R. M. Talbot, L. M. Hartley, K. Marzetta, and B. S. Wee, Transforming undergraduate science education with learning assistants: Student satisfaction in large-enrollment courses, *J. Coll. Sci. Teach.* **44**, 28 (2015).
- [40] R. M. Goertzen, E. Brewster, L. H. Kramer, L. Wells, and D. Jones, Moving toward change: Institutionalizing reform through implementation of the Learning Assistant model and Open Source Tutorials, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020105 (2011).
- [41] laa.colorado.edu.
- [42] V. K. Otero, S. J. Pollock, and N. D. Finkelstein, A physics department's role in preparing physics teachers: The Colorado learning assistant model, *Am. J. Phys.* **78**, 1218 (2010).
- [43] S. J. Pollock and N. Finkelstein, Impacts of curricular change: Implications from 8 years of data in introductory physics, *AIP Conf. Proc.* **1513**, 310 (2013).
- [44] J. K. Rice, *Teacher Quality: Understanding the Effectiveness of Teacher Attributes* (Economic Policy Institute, Washington, DC, 2003).
- [45] D. Sawada, M. D. Piburn, E. Judson, J. Turley, K. Falconer, R. Benford, and I. Bloom, Measuring reform practices in science and mathematics classrooms: the reformed teaching observation protocol, *School Sci. Math.* **102**, 245 (2002).
- [46] D. MacIsaac and K. Falconer, Reforming physics instruction via RTOP, *Phys. Teach.* **40**, 479 (2002).
- [47] M. D. Piburn and D. Sawada, Reformed Teaching Observation Protocol (RTOP): Reference Manual, ACEPT Technical Report No. IN00-3, Tempe, 2000.
- [48] www.learner.org.
- [49] A. Lakshmanan, B. P. Heath, A. Perlmutter, and M. Elder, The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction, *J. Res. Sci. Teach.* **48**, 534 (2011).
- [50] K. Ryker and D. McConnell, Can graduate teaching assistants teach inquiry-based geology labs effectively?, *J. Coll. Sci. Teach.* **44**, 56 (2014).
- [51] D. Ebert-May, T. L. Derting, T. P. Henkel, J. Middlemis Maher, J. L. Momsen, B. Arnold, and H. A. Passmore, Breaking the cycle: future faculty begin teaching with learner-centered strategies after professional development, *CBE Life Sci. Educ.* **14**, ar22 (2015).
- [52] T. Harms and R. M. Clifford, Assessing preschool environments with the early childhood environment rating scale, *Stud. Educ. Eval.* **8**, 261 (1982).
- [53] R. C. Pianta, J. Belsky, N. Vandergrift, R. Houts, and F. J. Morrison, Classroom effects on children's achievement trajectories in elementary school, *Am. Educ. Res. J.* **45**, 365 (2008).
- [54] H. C. Hill, L. Kapitula, and K. Umland, A validity argument approach to evaluating teacher value-added scores, *Am. Educ. Res. J.* **48**, 794 (2011).
- [55] D. M. Levin, D. Hammer, and J. E. Coffey, Novice teachers' attention to student thinking, *J. Teach. Educ.* **60**, 142 (2009).