Learning from Inquiry in Practice
ASP Conference Series, Vol. 436
Lisa Hunter and Anne J. Metevier, eds.

Improvements in Professional Development Program Participants' Understandings about Inclusive Teaching

Anne J. Metevier,[1,2,3] Lisa Hunter,[1,2,3,a] Barbara K. Goza,[4]
Lynne M. Raschke,[3,5] and Scott Seagroves[1,2,3]

[1]*Institute for Scientist & Engineer Educators, University of California, Santa Cruz, CA 95064*

[2]*Akamai Workforce Initiative, University of Hawai'i Institute for Astronomy, Advanced Technology Research Center, Pukalani (Maui) HI 96768*

[3]*Center for Adaptive Optics, University of California, Santa Cruz, CA 95064*

[4]*Department of Psychology, University of California, Santa Cruz, CA 95064*

[5]*School of Sciences, College of St. Scholastica, Duluth, MN 55811*

[a]*Project director*

Abstract. A major emphasis of the Center for Adaptive Optics Professional Development Program (PDP) is training early-career scientists and engineers to teach more inclusively as well as more effectively. To this end, the PDP includes workshops on diversity and equity, and PDP participants are explicitly encouraged to weave inclusive instructional strategies into the inquiry laboratory activities they design and teach. In an initial effort to gauge the effectiveness of the PDP's diversity and equity training, we have analyzed 2008 and 2009 PDP participants' responses to a survey knowledge question that asks them to briefly describe how they would engage a diverse undergraduate student population through their teaching and research. Each participant answered the survey question before any PDP training, as well as after a series of intensive PDP workshops. We developed a rubric to score and analyze participants' pre- and post-workshop responses, and have found that their response scores improve significantly after PDP training. This indicates that PDP training does improve participants' understandings about how to teach inclusively. Furthermore, survey respondents who participated in the PDP in both 2008 and 2009 showed little decrease in response scores between years, but continued increases with continued training. In this paper, we detail our rubric development, survey response scoring, analysis, and results, as well as the implications our results have had for refining our goals for PDP participants and for further improving PDP workshops.

1. Introduction

The Professional Development Program (PDP; Hunter et al. 2008 and Hunter et al., opening paper in this volume) is at the heart of a suite of education programs developed through the Center for Adaptive Optics (CfAO) and now continuing at the Institute

for Scientist & Engineer Educators (ISEE)[1] and Akamai Workforce Initiative (AWI)[2]. Major goals of these programs include training early-career academic scientists and engineers to teach more inclusively and effectively at the post-secondary level, and broadening participation of undergraduates in STEM (science, technology, engineering, and mathematics) while improving their preparation. Through the PDP, graduate students and postdoctoral researchers in STEM fields are trained to teach using inquiry methods, and are explicitly instructed to consider diversity and equity issues as they design and teach their own inquiry activities.

While many assume that the demographics have improved in the past few decades, encouraging students of all backgrounds to pursue STEM degrees and careers is still of vital importance to these fields. Roughly 30% of students in the U. S. aim to pursue STEM degrees upon entering college, no matter their racial or ethnic background. However, according to Hurtado, Eagan, & Chang (2010), only ∼20% of STEM majors from underrepresented minority (URM[3]) backgrounds achieve STEM bachelor's degrees within five years. The completion rate is significantly higher (30-45%) for White and Asian American STEM majors, though the average five-year rate of degree completion, regardless of students' background or their field of study, is ∼65%. The Hurtado et al. results imply that STEM majors struggle to complete their intended degrees, and that this affects URM students most strongly. The difficulties faced by URM STEM majors include discouraging effects such as "stereotype threat" and "disidentification". These are areas of intense recent current study; see, e.g., Steele & Aronson (1995) and Osborne (1995), respectively.

On another front, inquiry has been prominent in U. S. educational standards (NRC 1996, 2000) and hailed in both U. S. and international reform efforts (e.g., Osborne & Dillon 2008) as a way of teaching and learning STEM that improves students' abilities and interest in these fields. Research shows that inquiry learning can also have a positive effect on diversity and equity in STEM fields. For example, Wilson et al. (2010) find that high school science students receiving inquiry instruction reach a higher level of achievement than students taught using "commonplace" instruction methods. Furthermore, Wilson et al. measure lower achievement by non-White students (vs. White students) who are taught with commonplace methods, but they do not find a detectable achievement gap by race for students taught via inquiry.

Professional Development Program efforts to train scientists and engineers a) to teach more effectively through inquiry and b) to address diversity in the classroom are therefore clearly intertwined. Support for an explicit focus on diversity and equity in the PDP comes not only from the research presented above, but also from National Science Foundation funding of our program, which was awarded both for the intellectual merit of our work and for potential "broader impacts". To this end, the PDP staff and community project a philosophy that broadening participation in STEM is not merely a recruitment issue, but that there are strategies we can employ directly in the classroom, with the students we already have, to create an inclusive environment for learning and

---

[1]See ISEE website at `http://isee.ucsc.edu`

[2]See AWI website at `http://kopiko.ifa.hawaii.edu/akamai/`

[3]In their paper, Hurtado et al. (2010) define URM to mean underrepresented racial minorities, including Blacks, Latinos, and Native Americans.

to encourage all our students' success. In the PDP, we focus on diversity with regard to race, ethnicity and gender (note that Hill, Corbett, & St. Rose 2010 report that men still earn more bachelor's degrees than women in nearly every science and engineering field, with women earning as few as ∼20% of degrees in physics, computer science, and many engineering fields).

As part of their training, PDP participants engage in workshops on diversity and equity in which they review relevant demographic information and discuss topics such as stereotype threat and how to project a "malleable mindset" (Dweck 2006), the idea that one's intelligence can grow and is not fixed. Participants then design and teach their own inquiry activities. They are supported in making inclusive instructional choices through input from PDP staff and through an activity "design template" in which they must articulate the rationale behind each of their activity components. The CfAO and now ISEE and AWI educational programs have intentionally created and plugged into existing teaching venues that reach students of backgrounds that are not currently well represented in STEM fields. PDP participants teach their activities in these venues, and therefore gain valuable experience teaching students of many backgrounds.

The emphases of the PDP diversity-related training and a more complete description of that training are described in Hunter et al. (second paper in this volume). In this companion paper, we give a preliminary report on the effectiveness of the diversity-related training in the PDP workshops, which we have measured using data from pre- and post-workshop surveys administered in 2008 and 2009. In the next section of the paper, we discuss our methodology, giving background on the surveys and how we scored participants' responses. We present our results, which indicate significant knowledge gains, in §3. Finally, in §4, we describe how this analysis has informed our goals for PDP participants and helped us refine diversity-related workshops in the 2010 PDP cycle.

2. Methods

2.1. Participants, Surveys, and Initial Analysis Efforts

The year-long PDP cycle of activities begins with an intensive series of training workshops in which participants learn about inquiry teaching strategies, techniques for assessing students' learning, and diversity and equity considerations for the classroom. Participants then design and teach their own inquiry activities in ISEE and/or AWI affiliated venues and reflect on the experience. Most PDP participants are science and engineering graduate students, though many postdoctoral researchers and often a few faculty members also participate in the program. The PDP has a layered structure so that participants can attend more than one PDP cycle, learning something new about inquiry, assessment, and/or diversity and equity each year. The PDP complements participants' science and engineering research preparation with training in inclusive and effective teaching, which they can carry forward as they move toward future academic positions. Of course, our participants are self-selected in that they already value teaching, and many recognize the importance of addressing diversity in the classroom before entering the PDP. We value their input on the program and constantly evaluate various aspects of the PDP to discern whether we are meeting our goals for participants and ex-

panding their understandings about effective and inclusive teaching. Participants' input pushes us to evaluate and expand our own understandings, as well.

As part of our evaluation of PDP training activities, we survey participants at three points in the PDP cycle: before participation in any PDP activities (pre-workshop), after the main series of training workshops (post-workshop), and after participants teach and reflect on their experience (post-PDP). The surveys include self-efficacy ratings, ratings of perceived accomplishments through the PDP, ratings of the value of specific workshops and the PDP as a whole, and prompts for open-ended responses to questions regarding knowledge we hope participants will gain through their PDP involvement. Multi-year participants are required to complete the surveys each year. In this paper, we concentrate on measuring diversity- and equity-related knowledge gains from participation in the PDP workshops, so we draw our data from the pre- and post-workshop surveys but do not discuss the post-PDP surveys. More specifically, we report on our analysis of responses to an open-ended prompt that is given in both the pre- and post-workshop surveys:

> **You are applying for a faculty position**, and are asked how you will engage a diverse undergraduate student population through your teaching and in your research. Please give at least 3 bulleted points that you might include (include just the short bulleted points, we assume your actual response would be in the form of a carefully written teaching statement).

To begin our analysis, we generated a list of 20 categories or themes that we thought might come up in participants' responses. These themes included "teaching with exciting components", "varying the types of instruction", and "developing and supporting a collaborative classroom structure". We then read through a subset of 10 responses from the surveys administered in 2007. A colleague had entered these responses into a spreadsheet so that we were blind to the respondents' names and whether the responses were "pre" or "post". We note that the 2007 pre-workshop surveys were given to participants after they had already received PDP-related reading assignments, so some participants' pre-workshop responses are influenced by knowledge gained from the readings. In this sense, the 2007 survey responses are somewhat "tainted", but they serve as a useful training set for our analysis. In subsequent years, participants have been required to complete their pre-workshop surveys before receiving any readings, so their pre-workshop survey responses are more authentic gauges of their understandings before participating in the PDP.

Four of us categorized the subset of 10 responses according to the 20 themes we had generated, and then we compared our results. We found that several challenges came up at this point. For one, we were unsure how to resolve the fact that different raters categorized given responses differently. We also found that it was difficult to avoid making inferences or adding interpretation to responses while categorizing them. Lastly, although our list of categories accurately reflected many of the themes we found in participants' responses, these categories did not necessarily reflect the concepts we hoped participants would learn from their diversity-related PDP training. For example, the category "teaching with exciting components" was a common theme in responses, but did not capture the fact that we hoped PDP participants would shift their focus from what *they* thought was exciting toward engaging *their students'* interests. We decided that the latter was the sort of change we wanted to be able to measure.

Ultimately, we did not proceed with our initial method of categorizing survey responses. However, we mention it here because it helped us to clarify what a "good" response might look like, and what it might include. This early analysis helped us move toward more directly evaluating whether or not the PDP diversity and equity workshops were imparting the knowledge we had intended participants to learn, in part by helping us more clearly articulate our diversity-related goals for participants.

2.2. Rubric Development and Scoring

Throughout the many years that the PDP has been held, we have collaborated with several assessment and evaluation experts (see Goza et al., this volume, for more on related social science activities) who have advocated using rubrics to measure knowledge gains. In fact, we encourage and support our own participants in using rubrics to assess their own students. Therefore, as we considered how to move forward with our diversity-related evaluation effort, we decided to pursue a similar effort: We designed a rubric. We began by identifying the five most important areas (or "constructs") in which we hoped our participants would demonstrate or move toward positive understandings. These constructs are:

1. The way we teach *directly in the classroom and/or research environment* can have a positive impact on diversity and equity in STEM.
2. Instructors can draw from the interests, experiences, and backgrounds of their students in order to engage them.
3. Instructors can vary their teaching and assessment strategies in order to provide effective learning experiences for students who learn in different ways.
4. Teachers' and learners' assumptions about one another can affect learners and learning.
5. Creating a collaborative classroom environment can support students of all backgrounds.

We note that the first construct above is much broader than the rest, and is meant to convey a fundamental idea that underlies all of our diversity-related efforts through the PDP. We consider diversity and equity in STEM an issue that can be addressed directly in the classroom; it is not merely a recruitment issue, or an issue that students should only seek outside mentoring and support for. Instead, we want PDP participants to know that they can have a positive impact on all their students, in large part by considering constructs 2–5 as they design and teach laboratory activities.

In Tables 1 & 2, we show our final rubric, based on these constructs. To fill out the rubric entries, we considered what we would look for in participants' survey responses that would indicate they had met our expectations with regard to a given construct. We also considered how they might exceed our expectations with their responses, or what type of response would indicate that a participant had not met our expectations. We allowed for the possibility that a participant might not write anything about a given construct (a "neutral" response), as well. Categorizing possible responses in this way later helped us avoid adding interpretation to what participants had written. We attached

a number to each column of the rubric: "PDP expectation not met" (0), "neutral" (1), "PDP expectation met" (2), "exceeds PDP expectation" (3). We used these numbers to score responses.

Although we show our final rubric in Tables 1 & 2, we note here that we tested the rubric and iterated through several drafts before arriving at the final version shown. To test a draft of the rubric, two raters (LH and AJM) independently scored a subset of responses from the 2007 "training" set described above, and then compared scores. These scoring comparisons helped us articulate our expectations for PDP participants with respect to each construct, and thus helped us refine each element of the rubric. (Here we use the term "element" to refer to the descriptions of each of the five constructs and of each of the four levels of quality within each construct. In other words, each blurb in Tables 1 & 2 is considered a rubric element.) When we felt we had a near-final draft of the rubric ready, we conferred with an evaluation expert (BKG), who led us in a training session in which we made final adjustments to the rubric, discussed examples of each rubric element, and scored a subset of the 2008 survey responses. We reviewed these scores in order to ensure that we were interpreting the rubric in the same way as we approached the rest of the scoring.

Two raters (AJM and LH) proceeded to separately score the 2008 and 2009 survey responses. Again, colleagues had entered the responses into a spreadsheet so that we were blind to the names of the responding participants and did not know whether a given response was from a pre- or post-workshop survey. For a given survey response, each rater added the scores from the five rubric constructs for a "summed" response score that could range from 0 to 15. We then compared our summed scores; if they differed by more than 2, we discussed our scoring and came to "within 2" consensus. Our first-pass inter-rater reliability, measured using Pearson's correlation coefficient, was $r = 0.78$ for the 2008 response scores, and $r = 0.82$ for the 2009 response scores. In both cases, this measure improved to $r > 0.90$ after we came to "within 2" consensus, and in all cases (first and second passes, both years' data) these correlations are statistically significant with $p < 0.01$, or a less than 1% chance of finding such a strong correlation if the null hypothesis were true. While our first-pass scores were probably sufficient for this analysis, we felt more comfortable with the accuracy of our scores after reaching near-consensus. In the analysis in §3, we use the mean of the two raters' "within 2" consensus scores for each response.

In Table 3, we give examples of responses that "met" PDP expectations for each construct. When scoring, raters looked for thoughtful responses that indicated participants' understanding of diversity and equity issues and/or inclusive teaching strategies. If a participant used relevant terminology in their response (e.g., "avoid stereotype threat"), but did not provide further narrative indicating that s/he knew what the terminology meant, the response scored well (usually given a 2) for the pertinent construct(s), but did not receive the highest score (3). Often, a participant would describe how s/he might implement an inclusive teaching strategy, or would provide a motivation for using inclusive teaching strategies. These responses also scored well (again, usually given a 2 for the relevant constructs). A response that included both a motivation and a description of how to apply an inclusive strategy was generally given the highest score.

Table 1. First three constructs of rubric used to score PDP participants' responses to diversity- and equity-related survey prompt.

	PDP expectation *not met* (0)	**Neutral response (1)**	**PDP expectation met (2)**	**Response *exceeds* PDP expectation (3)**
1. Teaching can impact diversity and equity	Response shows no indication that one's own teaching in the classroom and/or research environment could make a difference. Response may hint at recruitment or other supports *outside* the classroom or lab.	Teaching in classroom and/or research environment is mentioned, but no direct link is made to impact on diversity and equity.	Response directly links teaching in classroom and/or research environment to students, effect on diversity and equity.	Response not only links teaching and effect on diversity and equity, but also elaborates a rationale (e.g., *why* a given teaching strategy might positively affect diversity and equity).
2. Consideration for learners' backgrounds, interests, experiences	Response focuses on what *teacher* wants to teach or thinks is interesting/exciting.	Response does not explicitly indicate consideration for learners backgrounds, interests, and/or experiences.	Response indicates a focus on learners' backgrounds, interests, and/or experiences.	Response focuses on learners' backgrounds, interests, and/or experiences, *and a reasoning or strategy* for engaging students' backgrounds/interests/ experiences.
3. Varying teaching and learning	Response tied to few teaching methods (e.g., lecture only) without regard for students' different ways of learning.	Response may mention many teaching methods, but does not explicitly link them to different ways of learning.	Response explicitly links varying teaching and learning to engage students with different ways of learning.	Response explicitly links varying teaching and learning to engage students with different ways of learning, and *elaborates on strategies, reasoning*.

Table 2. PDP diversity and equity rubric, continued.

	PDP expectation *not met* (0)	**Neutral response (1)**	**PDP expectation met (2)**	**Response *exceeds* PDP expectation (3)**
4. Teachers' and/or learners' assumptions or expectations about students	Response indicates an assumption/expectation that minority students are lower achieving/ability.	Response does not explicitly indicate assumptions about or expectations for students.	Response indicates an interest in addressing assumptions/expectations about students. May briefly mention "safe environment", "stereotype threat", "identity safety".	Response indicates an understanding of how expectations/assumptions can affect learners; and the need to maintain high expectations for all. May not mention buzz-phrases, but understanding of concepts behind these phrases is clear.
5. Classroom and/or research culture/ environment	Response is tied to teacher's viewpoint, does not indicate an interest in fostering a collaborative classroom and/or research culture/environment.	Response may indicate collaborative teaching strategies, but does not explicitly link to creating a collaborative classroom and/or research culture/environment.	Response links collaborative strategies to an intent to create a positive, inclusive classroom and/or research culture/environment.	Response links inclusive/collaborative strategies to an intent to create a positive classroom and/or research culture/environment, *and elaborates reasoning*.

Table 3. For each construct, we provide a sample participant response that was scored by both raters as having "met" the PDP expectation. Note that some examples have been edited so that they only include the portion of the response that is most relevant to the given construct.

Construct	Sample participant responses meeting PDP expectations (score = 2)
1. Teaching can impact diversity and equity	"Engage a varied set of learner assumptions and cultural sensitivities by building a participative and supportive classroom environment. Motivate participation by providing multiple entry points to content and support for varied learning styles. Teach science as science is actually done in order to motivate and retain a wide variety of learners."
2. Consideration for learners' backgrounds, interests, experiences	"Incorporate prior knowledge of students, work from that rather than starting from scratch. Include topics relevant to them: research topics they pick, projects they design (with help), etc."
3. Varying teaching and learning	"In teaching I will utilize a diverse set of styles of learning (inquiry, lecture, group work, labs, reports, etc. . .) to engage diverse students."
4. Teachers' and/or learners' assumptions or expectations about students	"Competence and achievement is expected from all students. There are clear learning objectives for all students, yet is there a differentiation in instructions to meet individual needs."
5. Classroom and/or research culture/ environment	"Make sure the classroom setting is safe and comfortable so that all the students are willing to participate. Group students in different ways throughout the semester so that they are comfortable working with as many students as possible."

3. Results

3.1. Knowledge Gains Demonstrated by "Pre" and "Post" Scores

In total, we scored 98 pre- and post-workshop responses from the 2008 and 2009 surveys; 44 of these response pairs were from 2008 PDP participants, and 54 were from 2009 participants. In 2008, 24 participants were new to the program, and 20 were returning; in 2009, 29 participants were new and 25 returned. Twenty-one of the returning participants in 2009 also participated in 2008, so the 98 response pairs we scored correspond to 77 individuals. Of the 21 people who participated in both 2008 and 2009, 10 were new in 2008, and 11 were already returners in 2008. Four of the returning participants in 2009 did not participate in 2008 but had participated in the PDP in earlier years.

In Table 4, we show the average pre- and post-workshop response scores and the standard deviations of those scores, separated by year. Individual response scores ranged from 1.0 to 12.5. One can see that the scores for all participants, in both years, increased. Using a paired samples t test (in which pre- and post-workshop responses for a given participant were "paired" in the analysis), we find that the participants' gains are statistically significant: we find $p < 0.01$, or a less than 1% chance of finding such a strong increase if the null hypothesis were true, for the 2008 sample and also for the 2009 sample.

Table 4. Pre- and post-workshop response scores for 2008 and 2009.

		Pre-workshop		Post-workshop	
responses	N	Avg. score	Std. Dev.	Avg. score	Std. Dev.
2008 all	44	5.35	3.03	6.58	2.55
2008 new	24	4.52	2.99	6.17	2.81
2008 returning	20	6.35	2.84	7.08	2.17
2009 all	54	4.93	2.50	6.35	2.41
2009 new	29	4.38	2.53	6.31	2.60
2009 returning	25	5.56	2.36	6.40	2.22

If we further divide the new and returning participants' responses, we find that the new participants showed a larger increase in average score than returners, indicating a larger knowledge gain. This is probably because new participants come to the PDP with less prior knowledge about the constructs we emphasize in the program. In 2008, new participants on average scored slightly below the "neutral" score of 5.0 pre-workshop. Through analysis of the variance of the scores, with new versus returning as the independent variable, we find that the new participants' average pre-workshop score was significantly lower ($p < 0.05$) than the returners' average pre-workshop score in 2008. The difference between new and returning participants' average pre-workshop scores was also marginally significant ($p < 0.10$) in 2009. In both years, however, there was no significant difference between new participants' and returners' average post-workshop scores. The scores in Table 4 indicate that participants who have gone through the PDP at least once generally meet our expectations for one or two diversity-related constructs. This can be inferred from new participants' "post" scores, and returning participants "pre" and "post" scores, which average around 6.0–7.0.

It is encouraging to see that, on average, PDP participants demonstrate statistically significant knowledge gains about our diversity-related constructs. However, one might argue that participants' knowledge gains do not seem particularly *large*. We note that there are some difficulties with looking at participants' knowledge gains only by analyzing their total pre- and post-workshop scores, summing across constructs. One potential issue is that the first construct on our rubric, the idea that teaching can impact diversity and equity, is a much broader concept than those covered by the other four constructs. In the future, we may weight the scores for the first construct differently, or we may separate out the analysis for that construct (see more on this in the Appendix). We also note that the average scores in Table 4 do not necessarily convey how large we think the knowledge gains were. To us, the PDP staff and researchers, there is not the same knowledge gain "distance" between scores of 0, 1, 2, and 3 for a given construct

on the rubric. Of most importance to us is movement from 0/1 ("PDP expectation not met"/"neutral response") to 2/3 (PDP expectation "met" or "exceeded"). Lastly, we point out that on the survey, participants were prompted to list three bullet points about engaging a diverse group of students, but we have rated them based on five constructs. Keeping this in mind, movement from a pre-workshop score of 5.0 (neutral on average) to a post-workshop score of 8.0 (expectations for three constructs met, others neutral) is excellent.

3.2. Knowledge Gains from Analysis of PDP Expectations Met/Exceeded

Another way to look at participants' diversity-related knowledge gains is to measure how many constructs participants' responses received a score of 2 or 3 for, indicating that they met or exceeded PDP expectations. In this analysis, we look more closely at response scores for individual constructs, rather than total summed response scores. For a given construct, we divide between scores of 1.0 or less, which indicate that participants' responses did not meet PDP expectations for that construct, and 1.5 or greater, which indicate that at least one of our two raters scored the response as having met PDP expectations. In Table 5, we show some results from this analysis. Here, we have separated out the responses that received a "met" score for two or more constructs versus those that did not.

Table 5. Number of constructs for which participants' responses met or exceeded PDP expectations.

	Pre-workshop		Post-workshop	
2008+09 responses	0 or 1 constructs met expectations	2 or more constructs met expectations	0 or 1 constructs met expectations	2 or more constructs met expectations
all (98)	67 (68%)	31 (32%)	33 (34%)	65 (66%)
new (53)	42 (79%)	11 (21%)	19 (36%)	34 (64%)
returning (45)	25 (56%)	20 (44%)	14 (31%)	31 (69%)

Pre-workshop, only one-third (32%) of participants' responses met PDP expectations for two or more constructs, whereas nearly two-thirds (66%) of responses met expectations for two or more constructs post-workshop. These numbers appear to be dominated by knowledge gains demonstrated by new participants, though returners show gains, as well. In Figure 1, we provide histograms that further break down the total number of constructs for which participants' responses met or exceeded PDP expectations. It is particularly striking that more than 50 out of 98 pre-workshop responses met expectations for no constructs. Again, this number is dominated by new participants' responses. Post-workshop, 77 of 98 responses meet PDP expectations for at least one construct. In general, these numbers indicate clear diversity- related knowledge gains made by both new and returning PDP participants, and some knowledge retention by returners.

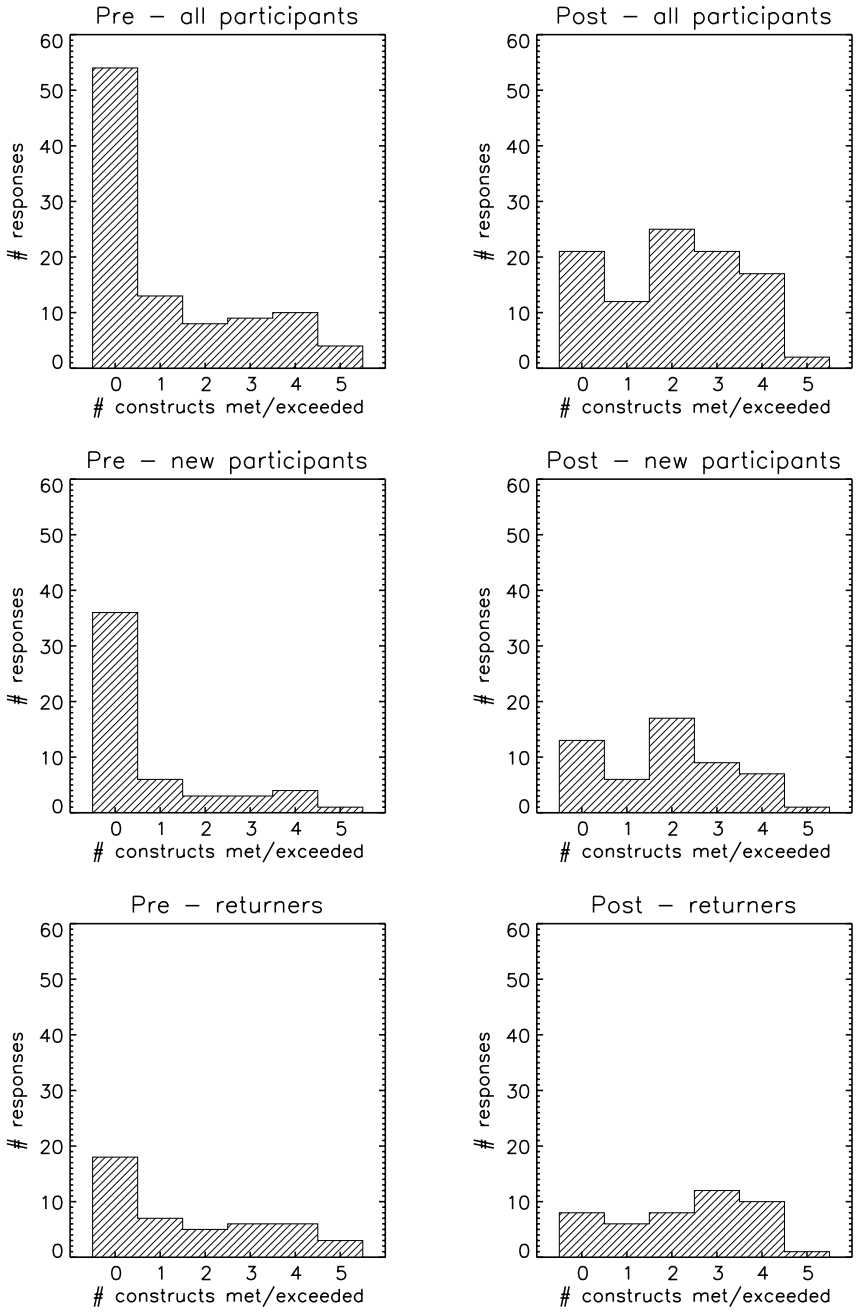

Figure 1. A more detailed view of the number of constructs for which participants' responses met or exceeded PDP expectations. To the left, we show pre-workshop responses, and post-workshop responses are on the right. All responses from 2008 and 2009 are shown in the top two panels, new participants' responses from both years are shown in the middle two panels, and returners' responses are shown at the bottom.

3.3. Indications of Returning Participants' Knowledge Retention

In the analysis shown in §3.1 and §3.2, we can see some hints of knowledge retention from the fact that returning participants generally have higher pre-workshop response scores than new participants. In this section, we look more closely at indications of diversity-related knowledge retention from the 21 participants who attended the PDP workshops in both 2008 and 2009. In Table 6, we show these participants' average pre- and post-workshop response scores from both years. A linear model of the four average scores (2008 "pre", 2008 "post", 2009 "pre", 2009 "post") for all 21 participants shows little statistical significance between the scores, only a marginally significant difference ($p < 0.10$) between the average 2008 pre-workshop response score and the average 2009 post-workshop response score. We acknowledge that the small number of subjects in this analysis makes it difficult to detect significant results. However, it is worth looking at the numbers more closely, as they provide practical implications that we can build upon as we analyze participants' responses in future years.

Table 6. Average response scores and standard deviations on those scores are shown for participants who attended both the 2008 and 2009 PDP series of workshops.

		2008				2009			
participants/ responses	N	Avg. pre	Std. Dev. pre	Avg. post	Std. Dev. post	Avg. pre	Std. Dev. pre	Avg. post	Std. Dev. post
all in both 2008 and 2009	21	4.79	2.88	5.95	2.46	5.52	2.42	6.24	2.22
new in 2008, returned in 2009	11	4.00	2.18	5.55	2.52	4.22	1.15	5.27	2.35
returned in 2008, also in 2009	10	5.65	3.40	6.40	2.45	6.95	2.69	7.30	1.55

From Table 6, one can see that there is an increase in the participants' average post- versus pre-workshop response scores from both 2008 and 2009. Furthermore, there is only a small drop between the 21 participants' average post-workshop score from 2008 and their average pre-workshop score from 2009. These numbers suggest that there was little knowledge lost between PDP cycles, and there were continued knowledge gains in both cycles. We can probe this further if we divide between the 11 participants who were new in 2008 and the 10 participants who were already returners in 2008. Here, we see that the returners' numbers show a steady increase, even between PDP cycles. On the other hand, the average pre- and post-workshop response scores for participants who were new in 2008 were nearly the same in 2009 as in 2008. This hints at the value of returning to the PDP for several years, and in particular the value of gaining practical teaching experience through the PDP, as this experience can further cement

the knowledge gained through the PDP workshops. However, we again acknowledge the small number of responses involved in this analysis.

3.4. Considerations for and Limitations of this Study

Here we take some time to mention important considerations around this pilot study as well as some limitations of the study. For one, we reiterate that our rubric was designed to assess participants' understandings related to specific diversity- and equity-related constructs, or concepts, that we emphasize in the PDP. These constructs are informed by research and by the support we see PDP participants needing as they design inquiry activities and teach their own students. However, we do not claim that these constructs cover the entire domain of diversity and equity issues, nor do we claim that our rubric allows us to assess participants' understandings of diversity and equity in general.

Our study relies upon the assumption that our rubric validly measures knowledge gains related to our constructs. We have demonstrated the *reliability* of the rubric by comparing two rater's scores. Also, in Table 3 above, we showed some example survey responses illustrating the "reasonability" of our rubric. However, we have not done an extensive test of the rubric's *validity*. Although we generally felt comfortable with our rubric and analysis, one thing we might have changed, in retrospect, is the wording of the diversity- and equity-related survey prompt. In the prompt, we asked PDP participants to write about how they would engage a diverse population of students through their teaching and in their research. While we ultimately want participants to connect the way they teach laboratory activities in the classroom to the way they mentor students in the research environment, the mention of research in the survey prompt led to some responses that did not seem directly relevant to teaching and mentoring.

We also note that post-workshop survey response scores may be affected by "survey fatigue". Participants receive pre-workshop surveys as part of their registration for the PDP, and they are able to fill out the surveys on their own time. Post-workshop surveys are much longer and are given at the very end of an intensive series of workshops. We have occasionally heard participants mention that they filled out their post-workshop surveys very briefly because they were tired at that point. This type of fatigue may have caused some participants' post-workshop responses to score fairly low.

Lastly, although our results indicate that participants made gains in their diversity-related understandings due to their participation in the PDP workshops, we understand that this does not necessarily mean that participants apply their new knowledge when they teach. Still, they need to have knowledge of the relevant issues and teaching strategies that can mediate those issues before they can intentionally design activities and teach them inclusively. This study demonstrates that the PDP workshops are successful in imparting relevant knowledge. Furthermore, we suspect that the PDP teaching experience serves to further cement and may even expand participants' diversity and equity understandings. In the future, we will test this hypothesis by carrying out further studies in which we will look more closely at participants' activity designs and teaching practices.

4. Conclusions and Implications for Further Professional Development Efforts

Our analysis shows that PDP participants do demonstrate statistically significant knowledge gains about diversity- and equity-related issues due to their participation in the PDP workshops. Furthermore, returning participants' survey response scores hint that these knowledge gains may be lasting. However, we might have hoped that participants' knowledge gains would be *larger*. Our results point to something that we sensed but had no more than anecdotal proof of before this analysis: Although we have given PDP participants a strong background understanding of the diversity- and equity-related issues that may affect their students, we need to give participants more practical strategies for addressing diversity and equity through their teaching. In particular, we need to make the links between teaching via inquiry and teaching inclusively clearer.

Creating the rubric for this analysis helped us focus on the points that we, as PDP staff, felt were most important for participants to learn about related to diversity and equity. In other words, it helped us clarify our goals for participants. These points, or emphases, in turn informed the rubric constructs. In the 2010 PDP cycle, we have drawn from and further refined these emphases, and we have translated them into clearer expectations for PDP participants (see a more thorough discussion in Hunter et al., second paper in this volume). In refining our goals for participants, we have brought forward a new emphasis on identity and students' feelings about belonging in the scientific and engineering community. We have also dropped the first rubric construct, the idea that we can effect a change by including diversity and equity considerations in our approach to classroom teaching, as an emphasis and instead present it as an over-arching part of our PDP philosophy.

Our refined diversity- and equity-related emphases within the PDP can now be listed this way:

- Learners should be provided with multiple ways to learn, communicate, and succeed in their work.
- Learners' goals, interests, and values should be engaged and leveraged in the classroom.
- Learners and teachers develop beliefs about learning, achievement, and teaching, and teachers should support an expectation of student success.
- Learners should have opportunities to collaborate with one another and equal opportunities to participate in activities.
- Learners should be supported in gaining a sense of belonging in the science and engineering culture.

We have written up a new document on these emphases which was required reading for participants in the 2010 PDP cycle (again, see Hunter et al.'s more thorough discussion). The document includes a description of diversity-related activity design and teaching expectations and examples of how a laboratory activity could be designed or taught with consideration for the PDP diversity and equity emphases. We have also made more explicit references to these emphases when presenting the diversity-related PDP workshops this year, and we are now making stronger, more explicit connections

between inquiry and inclusiveness. We developed a new workshop session this year in which participants discussed how the PDP diversity and equity emphases are highlighted when we look at the design and instruction of the Light and Shadow inquiry activity (this is an activity that all PDP participants experience as learners during the workshops). As we score and analyze participants' responses to the 2010 PDP surveys, it will be very interesting to see whether the changes we have made to PDP readings and workshops result in larger knowledge gains for participants.

The data from a small number of participants whose responses are in both the 2008 and 2009 sets hint that new participants may retain and/or apply diversity and equity considerations less effectively than returning participants. We are eager to see if these hints persist with larger numbers of responses. We will then explore (with larger datasets as well as through participant feedback) whether this may be because new participants are overwhelmed by training in not only diversity/equity, but also in inquiry, assessment, and effective pedagogy more generally. Another possibility is that new participants make what might be called "awareness" gains but require a full PDP cycle (including teaching experience) before they are ready for deeper gains in understanding and application.

This evaluation effort has not only allowed us to demonstrate that our workshops improve participants' knowledge about diversity and equity, but has also helped us clarify and refine our expectations for participants. Within the PDP community, this effort has been an interdisciplinary collaboration between physical and social scientists who are also PDP staff educators and researchers. We emphasize that internal evaluation efforts like this one can be very effective in informing and refining program goals. We hope that our results demonstrate to the larger community of science and engineering educators that it is possible and worthwhile to build diversity and equity considerations into college-level curricula.

Acknowledgments. We are very grateful to Glenn Reyes, Ashley Hyde-Smith, Hilary O'Bryan, and Amanda Martinez for data entry. We also thank Jerome Shaw and the AScILS collaboration for teaching us about the utility of analyzing survey responses using rubrics. This material is based upon work supported by: the National Science Foundation (NSF) Science and Technology Center program through the Center for Adaptive Optics, managed by the University of California at Santa Cruz (UCSC) under cooperative agreement AST#9876783; NSF DUE#0816754; NSF AST#0710699; NSF AST#0836053; the Air Force Office of Scientific Research (via NSF AST#0710699); UCSC Institute for Scientist & Engineer Educators. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding sources.

References

Dweck, C. S. 2006, Mindset: The New Psychology of Success (New York: Random House). URL http://mindsetonline.com/

Hill, C., Corbett, C., & St. Rose, A. 2010, Why So Few? Women in Science, Technology, Engineering, and Mathematics, AAUW, Washington DC. URL http://www.aauw.org/learn/research/whysofew.cfm

Hunter, L., Metevier, A., Seagroves, S., Porter, J., Raschke, L., Kluger-Bell, B., Brown, C., Jonsson, P., & Ash, D. 2008, Cultivating Scientist- and Engineer-Educators: The CfAO

Professional Development Program. URL http://isee.ucsc.edu/participants/programs/CfAO_Prof_Dev_Program.pdf

Hurtado, S., Eagan, K., & Chang, M. 2010, Degrees of Success: Bachelor's Degree Completion Rates Among Initial STEM Majors (Higher Education Research Institute, UCLA). URL http://www.heri.ucla.edu/nih/HERI_ResearchBrief_OL_2010_STEM.pdf

NRC 1996, National Science Education Standards (Washington DC: National Academies Press). National Research Council: National Committee on Science Education Standards and Assessment, URL http://www.nap.edu/readingroom/books/nses/

— 2000, Inquiry and the National Science Education Standards (Washington DC: National Academies Press). National Research Council: Committee on the Development of an Addendum to the National Science Education Standards on Scientific Inquiry; editors S. Olson and S. Loucks-Horsley, URL http://fermat.nap.edu/html/inquiry_addendum/

Osborne, J. F., & Dillon, J. 2008, Science Education in Europe: Critical Reflections, A Report to the Nuffield Foundation. URL http://www.kcl.ac.uk/content/1/c6/01/32/03/SciEdinEuropeReportFinal2.pdf

Osborne, J. W. 1995, Personality and Social Psychology Bulletin, 21, 449

Steele, C. M., & Aronson, J. 1995, Journal of Personality and Social Psychology, 69, 797

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. 2010, Journal of Research in Science Teaching, 47, 276

Appendix A. Knowledge Gains Per Construct

In §3.1 above, we noted that the first construct of our rubric for assessing PDP participants' diversity- and equity-related knowledge gains is somewhat broader conceptually than the other constructs. We have considered separating out our analysis of response scores for the first construct, and in this vein, we show 2008 and 2009 PDP participants' average response scores per construct in Figure 2. In the top panel, which shows all participants' average response scores, this figure does indicate slightly larger knowledge gains related to the first construct (from an average pre-workshop score of 1.31 to an average post-workshop score of 1.66) than for other constructs.

From the workshop designers' perspective, it is helpful to look at the average response scores for each of the constructs. This way, we can see whether participants' understandings are improving in each of the diversity- and equity-related areas we have designated as being most important in the PDP. In Figure 2, one can see the very positive result that both new and returning participants made gains related to every construct. Participants seem to have made the weakest gains related to the third construct (on varying teaching and learning), which indicates that we could do more to bolster the training participants receive in this area. Other things we can see from the figure: new participants come to the PDP with less knowledge per construct than returners, but make strong gains. Returning participants' pre-workshop scores are higher than new participants', indicating that returners retain some knowledge in each construct area, and returners continue to gain from repeated participation in the PDP.

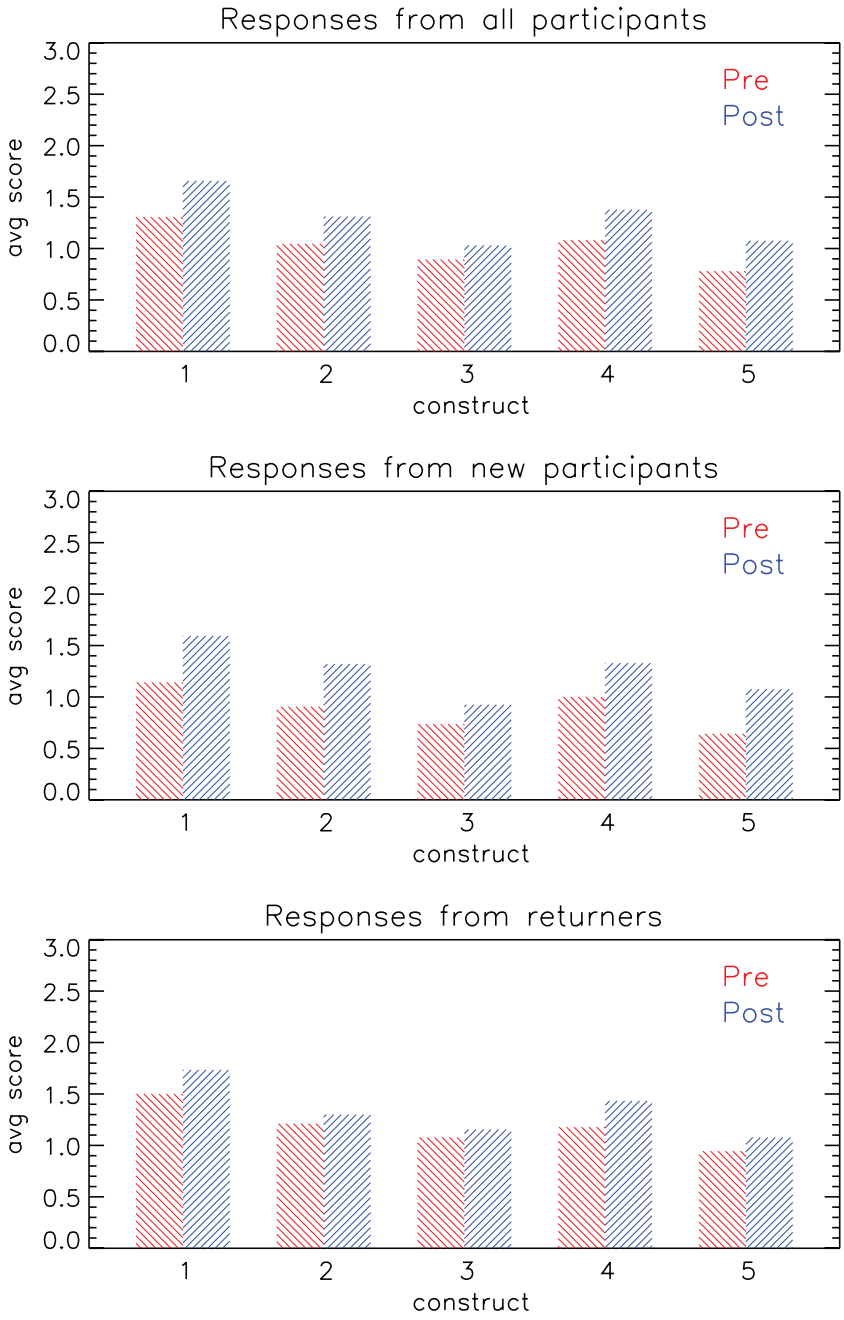

Figure 2. Average scores from 2008 and 2009 PDP participants' responses are shown for each construct. Red bars indicate average pre-workshop scores for a given construct; blue bars indicate average post-workshop scores. All participants' response scores are shown in the top panel, new participants' response scores are shown in the middle panel, and returning participants' scores are shown at the bottom.