Evaluation methodology and results for the new faculty workshops

Stephanie V. Chasteen¹, Rajendra Chattergoon², Edward Prather³, and Robert Hilborn⁴

¹*University of Colorado Boulder, UCB 390, Boulder CO 80309*

²*School of Education, University of Colorado Boulder, UCB 249, Boulder CO 80309*

³*Department of Astronomy, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721*

⁴*American Association of Physics Teachers, One Physics Ellipse, College Park, MD 20740*

This paper describes the current evaluation of the Physics and Astronomy New Faculty Workshop (NFW) as a case-study in evaluation of professional development workshops. We describe a Theory of Action (ToA) for the workshop, and the evaluation methods and measures. The evaluation suggests that the ToA of the workshop is only partially fulfilled: workshop experiences are positive, and participants gain knowledge of active learning, but participants have room for additional growth in skill, self-efficacy and social support in their use of active learning. We discuss the implications of these results for the NFW program and evaluation.

I. INTRODUCTION

The Physics and Astronomy New Faculty Workshop (NFW) has provided multi-day professional development (PD) workshops since 1996. The workshops are organized by several professional societies and include education research experts and practitioners as presenters. The NFW are aimed at physics and astronomy faculty within the first years of their initial teaching appointments. The mission of the NFW is to improve student learning by helping all faculty become knowledgeable users of evidence-based instructional practices (EBIPs), thus improving student learning and retention.

It has been almost a decade since the last publication on the NFW evaluation [1], which demonstrated positive effects on participant awareness of EBIPs. Much has changed in this time. Faculty are more aware of innovations in STEM education, and the STEM education research (SER) community now knows much more about what supports faculty success in changing their STEM education practice [2]. The NFW evaluation has evolved to address these best-practices. This paper focuses on the evaluation measurements used, including the analyses and interpretation, to help answer the question, “How can we provide high-quality evaluation of STEM faculty PD experiences?” We hope to build our capacity as a community to provide such meaningful evaluation, thus supporting broad change in STEM educational practice.

II. BACKGROUND AND THEORY

To achieve its mission, the NFW has three main goals: (1) Reach a large fraction of tenure-track faculty in physics and astronomy, (2) Help participants develop knowledge about recent developments in pedagogy, and (3) Help participants to integrate workshop ideas and materials in their classrooms. The workshop has been largely successful at achieving the first two goals: From 2008-2013 the NFW attendees represented 48% of tenure-track hires in physics and astronomy, and prior work has shown that the NFW has been successful in developing partici-

pants’ knowledge about, and attempt to use, EBIPs [1,2]. However, typical faculty knowledge of EBIPs can often be shallow, and many faculty discontinue use of EBIPs, and/or use them in ways that are not consistent with the original design [3]. Thus, using Rogers’ model of diffusion of innovations [4], the NFW has been successful at increasing knowledge, and persuading faculty to try to use EBIPs, but the effectiveness and sustainability of implementation is an ongoing challenge.

The purpose of this evaluation is to provide guidance to organizers as they plan the workshop, and to provide summative evaluation to stakeholders. Unlike research, the evaluation does not provide a validated study of research questions about the nature of professional development. Instead, the evaluation focuses on the question, “Do we observe the participant outcomes expected to derive from the NFW strategy?” First, we describe the underlying rationale for the NFW. This “Theory of Action” (ToA), describes the hypothesized connections between Workshop Inputs, Participant Outcomes, and Long-Term Objectives as follows, with measurable inputs/outcomes in *italics*: Through various **Workshop Inputs** of *information, access to experts, peer networking, and active engagement* [5] in a disciplinary setting, we can generate desired **Participant Outcomes**, including a sense of *competence, autonomy, and relatedness* (i.e. self-determination [6]) vis-à-vis the use of EBIPs, supporting **Long-Term Objectives** of *awareness, decision to use, and long-term effective use* of EBIPs [2, 4].

The current evaluation has been designed iteratively to address key questions regarding the workshop design and impacts, alignment with educational theory and findings from educational change research, and degree to which it respects the voices multiple stakeholders – i.e., participants, presenters, organizers, and funding agencies.

III. METHODS

We report on the most recent offerings of the NFW [7] in 2015 (June 22-25, N=75 and November 19-22, N=65),

and 2016 (June 20-23, N=52), though post-workshop data for the last cohort was not available at the writing of this paper. Participants are typically 70% male, primarily Caucasian (~60-70%) with a significant Asian representation (~20-30%). Many (~35%) received an undergraduate degree outside North America, about half of participants are employed at undergraduate-only and half at research-focused institutions. The median number of years of teaching experience is 2.

The evaluation was designed based on the ToA. Survey instruments were informed by former evaluation efforts of the NFW [1] and physics faculty [3] and similar efforts in mathematics [8] and chemistry [9]. Evaluation data comes from pre- and post-workshop surveys [10], field notes, and workshop observations, plus a new one-year follow-up survey to assess longer term impacts. Survey instruments have been heavily iterated based on data and discussion, and fine-tuned to improve validity, leveraging author RC's expertise in measurement. The June 2016 cohort received the final instruments (see [10]).

Our surveys use a set of items we call "Participant Outcome" measures. These seven Likert-scale questions, given pre-post, are intended to measure variables that could impact participants' ability to use active learning effectively: competence, autonomy, and relatedness [6]. We have operationalized these three factors as *knowledge*, *skill*, *motivation*, belief in the *effectiveness* of, *self-efficacy* (to support others, and get good student evaluations), and *social support* in using active learning (adapted from work by Hayward et al. [8]). Participant Outcomes items are analyzed for a pre/post matched sample. Note that we choose to report the median (and median gain) on these items to avoid misleading results from reporting the mean for these Likert scale questions: Since response categories are not an interval scale, and a respondent cannot answer between two scale scores, mean responses of "3.3" vs. "3.7," for example, are difficult to interpret meaningfully and may lead to identifying differences which are mainly artifacts of the numerical analysis. "Median gain" is defined as the median of individual gains. We note, however, that the median represents only the 50th percentile of the response distribution and does not take into account the distribution of responses. We argue this is a valid approach as it reduces the weight of outliers, and our sample responses tend to cluster around the mean and median. We also examine the (more traditionally reported) mean, standard deviation (SD) on the mean, mean gain, and effect size to ensure valid interpretation; effect sizes are computed by dividing the mean (not median) gain by the pooled standard deviation.

IV. RESULTS

Here we present selected measurements used to assess our Theory of Action (ToA), and describe how evaluation results can be used to provide recommendations.

A. Before: Participant characteristics

To fulfill our ToA, workshop sessions should respect and build on participants' prior beliefs and experience. Thus, the pre-workshop survey includes questions regarding demographic data, teaching experience, areas of interest, knowledge of common EBIPs, and a validated survey of instructional practice [11] (not reported here) in addition to the seven Participant Outcomes items.

There is some diversity among participants (see Methods), suggesting that differentiated instruction may be helpful: While most are in their first years of teaching, about 1/3 have more than 2 years of teaching experience, and about 1/3 of participants did not experience the U.S. educational system as an undergraduate. On the Participant Outcome items (see Table I), we find that participants have high incoming levels of belief in the effectiveness of active learning, and motivation to use it (both medians are at the scale maximum pre-workshop), and low levels of perceived skill. This has led us to recommend that presenters focus less on "convincing" participants to use EBIPs, where there is little room for growth, but instead focus on increasing participants' skill and confidence to use active learning (e.g., by giving concrete strategies and hands-on practice).

How familiar are participants with common EBIPs? Data are shown as a frequency graph (see Fig. 1), which enables easy comparison of levels of use and familiarity (or lack thereof) across EBIPs and Concept Inventories (CIs). Across workshop cohorts, respondents are most familiar with Peer Instruction, PhET, Cooperative Group Problem Solving, and CIs. About half of the June 2016 cohort has used at least one listed EBIP (though about half of those have discontinued their use of any given EBIP); slightly higher than previous evaluations [1].

FIG. 1. June 2016 EBIP familiarity (sample)

B. During: Workshop experiences

In the context of the NFW, Workshop Inputs include participant-centered sessions that build on existing faculty knowledge, use peer discussions, and address possible challenges in adoption [1-3]; these best-practices are encapsulated in a series of Presenter Tips [5]. Does the workshop appear to follow the recommended Workshop Inputs? To address this question, the post-workshop survey includes a series of questions asking participants to

rate each session and the workshop as a whole, including open ended questions (what was best, what could be improved) and Likert ratings. Responses across items, as well as workshop observations, are coordinated to provide rich feedback on the session organization. Common complaints in open-ended questions have been used to generate closed-choice questions to assess the extent of those concerns, and these are displayed in a frequency plot to allow easy interpretation (see Fig. 2 for a sample).

While the outcomes and feedback are largely glowingly positive (e.g., 91% agree that the overall quality of the workshop exceeded that of other PD workshops), constructive criticism has allowed organizers to make substantial modifications. For example, in June 2015, open-ended responses indicated exhaustion, limiting participants ability to engage effectively. The Nov. 2015 NFW addressed this issue through later start times and removing sessions; two closed-choice questions showed fewer complaints about cognitive overload (see Fig. 2). As another evaluation example: In June 2015, open-ended comments indicated that presenters were seen as giving too much of a “sales pitch” for the technique instead of practical implementation details. These results were communicated to organizers and a closed-choice question was created for Nov. 2015 identified fewer complaints.

FIG 2. Workshop ratings, Nov 2015 (sample)

Other findings are more nuanced. Survey results show a tension between providing a broad survey of techniques (cited as one of the most beneficial aspects of the NFW), and doing a “deeper dive” to give concrete skills in implementation (many request more implementation details). This can be interpreted as a satisfaction with the *knowledge* gained (as per our ToA) but a request for more *skill*. There is ambivalence across the cohort in terms of how to best achieve that skill: In November 2015, about half (53%) of participants indicated they would have liked more time to practice what they learned. As the workshop organizers add new sessions (such as working sessions giving time to practice creating and implementing

Peer Instruction questions, which were offered in Nov. 2015), we have been able to show that these sessions are quite popular, and are at least partially fulfilling the requested focus on skill and implementation.

C. After: Participant outcomes

The ultimate goal of the NFW is to support the effective use of EBIPs over time. The SER community is actively working on instruments to measure teaching practice. We give one such instrument [11] pre-workshop and one year later; results of this survey are forthcoming.

In the short-term, what we can assess are participants’ self-reported gains on our hypothesized antecedents to using active learning effectively: our Participant Outcome items. Results are shown below in Table I.

TABLE I. Pre-post scores on Participant Outcome items[10]. Effect size is computed from mean gain (see Methods). Sample sizes vary as not all cohorts received all items.

Item	n	Median (SD)		Median Gain (SD)	Effect Size
		Pre	Post		
Knowledge	88	3 (0.7)	3 (0.6)	0 (0.8)	0.7
Skill	89	2 (0.7)	3 (0.6)	0 (0.7)	0.4
Effectiveness	90	4 (1.0)	3 (0.7)	0 (1.0)	-0.3
Motivation	47	4 (0.8)	4 (0.5)	0 (0.8)	0.4
Social support	43	3 (0.8)	4 (0.8)	0 (0.8)	0.0
Support other	43	3 (1.0)	3 (0.8)	0 (1.1)	0.2
Evaluations	43	3 (1.3)	3 (1.0)	0 (1.4)	0.2

Though median responses on these items are typically higher post-workshop across the cohort as a whole, when we examine how much the scores of *individuals* increased as a whole (the median of the gain), we see no effect pre/post. This is borne out by examination of item means, which typically differ by less than 0.3 pre/post. However, the relatively large effect size for the mean gain on “knowledge” (mean 2.8 vs. 3.3 pre/post) suggests that the workshop does impact knowledge of active learning, as reported elsewhere [1,2]. The only negative shift is in participant beliefs in the effectiveness of active learning, but the effect size is small; the difference is essentially zero. We do see that SDs tend to decrease post-workshop, though this should not be over-interpreted (response categories are not an interval scale, and are likely hitting the response “ceiling”). The first four items (*knowledge* to *motivation*) are adapted from workshops on inquiry-based learning in mathematics [6]; in comparison, NFW participants show higher pre-workshop levels (with very little room for growth) but smaller gains, particularly in belief in the effectiveness of active learning.

Investigating results by participant demographics, we find no difference in the Participant Outcomes items by gender, but we see larger median gains in reported knowledge of active learning for those who did *not* receive an undergraduate degree in N. America (1.0) compared to those educated in N. America (0.0). Those with

more teaching experience (3 or more years) tend to score at a higher level on the Participant Outcomes measures (both pre and post), but their gains on these items are similar to more novice instructors; the same is true for those with high levels of belief in the effectiveness of active-learning pre-workshop. Thus, our evaluation measures provide evidence that the workshop is valuable for diverse faculty, but indicate an influence of incoming experience.

V. DISCUSSION AND FUTURE WORK

We have developed a Theory of Action of the NFW which posits that the workshop design will improve faculty's short-term perceptions of active learning and confidence in enacting it, thus leading to long-term, broad use of evidence-based teaching practice in the discipline. Our measures are able to provide evidence in support of the ToA, indicate the successes of the workshop, and suggest areas that can be improved. We claim that the evaluation design addresses the needs of multiple stakeholders, providing a voice for the participants, showing return on investment for the funding agency, as well giving guidance to the organizers. Multiple data sets, and concise visuals, are used to substantiate recommendations.

The main measurable impact is in increased *knowledge* of active learning strategies, and participants particularly comment that they benefit from the broad exposure to teaching techniques. Common critiques have begun to be mitigated, and the workshop is equally effective regardless of teaching experience or gender. We find that participants have high levels of incoming interest in using active learning, and that the NFW impacts their knowledge of EBIPs. Thus the *knowledge*, *belief in the effectiveness* and the *motivation* aspects of our ToA need less attention than *self-efficacy*, *skill*, and *social support*.

We find that participants are somewhat diverse, representing undergraduate-serving and R1 institutions, many have tried an EBIP, and some have substantial teaching experience. This diversity has led the evaluation to recommend differentiated instruction in the NFW, such as a "tracked" schedule, as well as more opportunities to share this experience through peer discussions.

Our ToA allows us to put common complaints into a theoretical framework: Comments that active learning is being "over-sold" may suggest that the NFW is erroneously focusing on *motivation*, which can sometimes prompt a negative response among participants, consistent with self-regulation theory [6]. Additionally, a presenter who is focused on "convincing" can be more cautious about highlighting challenges of EBIPs and aspects of successful implementation [1-3]. Given participants' interest in a broad survey of techniques, and their 50/50 split on needing more time to practice what they learned, the evaluation has recommended providing some sessions with an explicit focus on *knowledge* (e.g., broad survey sessions), but explicitly focus others on *skill* and *self-efficacy*, with a goal of encouraging participants to be more self-reflective teachers. Another workshop [8] demonstrated higher levels of gain in *skill*, and included hands-on activities, work time, and video observations. The NFW is in the process of incorporating more of such strategies, including online faculty learning communities [12], which may enable further gains in this area.

A main limitation of the measures is their self-report nature, though responses are quite consistent across workshops. Another limitation is the small room for growth on our Participant Outcome measures, limiting observable effects, and suggesting the need for alternative tools. Additional data, with our finalized instruments, will allow further investigation of participant outcomes and the influence of background variables. Use of the newly developed real time workshop observation protocol [13] may give additional insights into practice within individual workshop sessions, and how they support the ToA. Future evaluations will examine whether the workshop achieves long-term goals of impacts in reported teaching practice.

ACKNOWLEDGEMENTS

This work is supported by NSF DUE number 1431638 and Chasteen Educational Consulting. Special thanks to: Charles Hayward, Marilyne Stains, Charles Henderson, Alice Olmstead, Chandra Turpen, and the NFW team.

[1] C. Henderson, Am. J. Phys., **76**, 179 (2008).
 [2] C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj, Phys. Rev. ST – Phys. Ed. Res., **8**, 020104 (2012).
 [3] M. Dancy and C. Henderson, Am. J. Phys., **78** (10), 1056 (2010).
 [4] E. M. Rogers, *Diffusion of Innovations*, 4th ed. (Free Press, New York, 1995).
 [5] Presenter Tips: <http://bit.ly/1tsYoCc>
 [6] E. L. Deci & R. M. Ryan, *Psychological Inquiry*, **11**, 227 (2000).
 [7] <http://www.aapt.org/conferences/newfaculty/nfw.cfm>

[8] C. N. Hayward, M. Kogan, and S. Laursen, *Int. J. Res. Undergrad. Math. Ed.*, **2** (1), 59 (2016).
 [9] L.A. Baker, D. Chakraverty, L. Columbus, A.L. Feig, W. J. Jenks, M. Pilarz, M. Stains, R. Waterman and J. L. Wesemann, *J. Chem. Educ.*, **91**(11), 1874 (2014).
 [10] Sample surveys available at <http://bit.ly/NFWeval>
 [11] E. M. Walter, A. L. Beach, C. Henderson and C. T. Williams (submitted).
 [12] A. Rundquist, J. Corbo, S. Chasteen, M. Martinuk, C. Henderson and M. Dancy, PERC 2015, College Park, MD, 2015.
 [13] A. Olmstead and C. Turpen, PERC 2015, College Park, MD, 2015.