

2018

Transforming Undergraduate STEM Education: Reforming Classrooms by Training Current and Future Faculty

Karen E. Bailey

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Bailey, Karen E., "Transforming Undergraduate STEM Education: Reforming Classrooms by Training Current and Future Faculty" (2018). *Graduate Theses, Dissertations, and Problem Reports*. 5143.
<https://researchrepository.wvu.edu/etd/5143>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**Transforming undergraduate STEM education:
Reforming classrooms by training current and future faculty**

Karen E. Bailey

**Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Doctor of Philosophy
In
Biological Sciences**

**Michelle Withers Ph.D., Chair
Jonathan Cumming, Ph.D., Co-Chair
Michelle Richards-Babb, Ph.D.
Jim Belanger, Ph.D.
Katrina Stewart, Ph.D.**

**Department of Biology
Morgantown, West Virginia
2018**

Keywords: Scientific teaching, active learning, professional development, evidence-based practices, STEM, post-secondary, future faculty

Copyright 2018, Karen Bailey

ABSTRACT

Transforming undergraduate STEM education: Reforming classrooms by training current and future faculty

Karen E. Bailey

In the absence of formal pedagogical training, educators tend to teach the way they were taught, using traditional lecturing. Numerous studies have shown that engaging students in active learning improves student learning and retention of knowledge. Reforming STEM education has become a national imperative. Since the turn of the millennium, many initiatives have sprung forth in response to a growing awareness of the need for reform in post-secondary biology education. The National Academies Summer Institutes (NASI) is one such initiative. The Summer Institutes provide training in evidence-based strategies for biology educators from primarily research universities where lecture-based teaching in large enrollment introductory classes results in unacceptable rates of attrition. In 2010, the first regional offshoot of the SIs was piloted at West Virginia University (NASI@WVU) which expanded the reach of the SIs by serving current and future STEM faculty from any post-secondary institution. Analysis of exit surveys indicated that the original SI, participants of the NASI@WVU highly valued the experience and reported increases in teaching knowledge, skill and implementation of active learning strategies regardless of discipline or rank. While current and future faculty reported differences in their opportunities to teach and reported using active learning strategies at different frequencies, they both reported using active learning strategies at higher rates after the NASI@WVU compared to before. A small theme analysis of open-ended responses on support to implement scientific teaching or challenges they would face revealed that both current and future faculty identified peer support as key in improving their instruction. When teaching videos of participants were evaluated with validated observation protocols, the participants were found to use higher

levels of both active and student-centered strategies ($p > 0.05$) than a control cohort of STEM faculty who did not take part in the NASI@WVU. In addition to directly training current and future faculty, the SIs have provided a model for programs to train graduate students and teaching assistants (TAs) in evidence-based teaching. The program developed at WVU to train graduate TAs for the newly reformed introductory biology teaching laboratory (BIOL 115/117) employed a combination of SI workshops and peer-review to foster reflective teaching practices and adoption of student-centered approaches. Teaching assistants who took part in this training program were evaluated with two observation protocols. Teaching assistants were found not to be significantly higher in their use of active learning or student-centered strategies after two rounds of peer-review that focused feedback on increasing use of these strategies. Varying outcomes for the graduate teaching assistants indicate that while training in scientific teaching does create changes toward student-centeredness, further effort is required to produce improvements similar to those seen in faculty instructors.

I dedicate my dissertation to my mom, Myrna Bailey. You always believed in me and always told me that I was destined for great things. You may not have always been able to give me everything I asked for but you gave me something even more important; a love of education, a love for seeking out new information. It is this determination that you instilled in me that has pushed me even when I felt like giving up. You always tell me you are proud of me but I am proud of you for dedicating your life to your children's education even when you didn't have the same opportunity. Thank you! I love you.

ACKNOWLEDGMENTS

In 2008, I received an opportunity of a lifetime for a young science high school teacher from Belize; I was going to a city named Morgantown to spend four months working with a West Virginia University organization called HSTA (Health Sciences & Technology Academy). During my four month “sabbatical” in Morgantown, I met Michelle Withers. During our first conversation, I realized this educator was like me; she understood that science education is important and the current methods of instruction need reform if we are to see any improvement in STEM education.

Once I started graduate school in the Biology department, I fell in love with Morgantown and its lifestyle. During my time in Michelle’s research program, I have had the opportunity to delve into areas of science education research that I would not have had in my home country. Michelle has been my sounding board and constant cheerleader as I have navigated the various aspects of my graduate education. She not only opened her laboratory to me, but her home as well, and I consider Jim, Fred and Barney a part of my extended family.

To my lab mates, Bia, Robin and especially my sometimes partner in “crime”, Nick, you have made my time in the Biology department a memorable and enjoyable one.

I want to thank my committee, Drs. Cumming, Belanger, Richards-Babb and Stewart, for your patience as we pioneered our way through being the first education laboratory in the Biology Department,. A special thank you to Katrina Stewart who, aside from serving on my committee, has also been a wonderful laboratory manager for the many years that I have served as a teaching assistant.

To my Morgantown family, the Fetty/Andersons, who, after two weeks in Morgantown, took me and my 7 year old child, under their wings and have been my constants ever since, thank you. To Kaitlyn, you took mommy’s graduate education in stride. The many nights spent at Allen Hall just outside my classroom until 10pm so that I could go to class and be a mom at

the same time is beyond what is expected of any 8yr old but somehow you always told me that you understood and it was ok. To Aden and Bailey, our “lab babies” (as we call them in the Withers lab), I could not imagine you not being a part of this journey. To my father-in-law, Ernest, for all the jokes and countless homemade tortillas that made it easier to spend many hours doing research, thank you. To my brother Errol, for checking in on me at just the right time.

Finally, to my husband Alden, your quiet support and broad shoulders, that I have cried and laughed on too many times to count, helped to make this a reality. You are my superman. It takes a village of love and support to achieve a Ph.D. and my village is awesome.

TABLE OF CONTENTS

Section	Page(s)
Abstract	ii-iii
Dedication	iv
Acknowledgements	v-vi
Table of Contents	vii
List of Figures	viii
List of Tables	ix
List of validated surveys	x
Chapter 1: Introduction	1-4
Transforming Science Education	1-5
Chapter 2: Participation in an adaption of the National Academies Summer Institute...	6-31
Abstract	6
Introduction	7-9
Materials and Methods	9-12
Results	13-23
Discussion	23-27
Acknowledgements	27
References	28-31
Chapter 3: Professional development workshops do make a difference...	32-59
Abstract	32
Introduction	33-36
Methods	36-43
Results	43-51
Discussion	51-54
Acknowledgments	55
References	55-59
Chapter 4: Exposure to reformed science teaching and peer evaluation may improve Future Faculty teaching	60-87
Abstract	60
Introduction	61-63
Methods	64-67
Results	67-74
Discussion	74-77
References	77-79
Chapter 5: General Discussion	80-84
References	85-88

LIST OF FIGURES

Figures	Page
Chapter 2: Participation in an adaption of the National Academies Summer Institute...	
Fig. 1. Representation of attendees by discipline	14
Fig. 2. Perception of teaching skills differs before and after participation at the NASI@WVU	16
Fig. 3A. Reported implementation of Scientific Teaching Strategies – Average percent of respondents	21
Fig. 3B. Reported implementation of Scientific Teaching strategies – Average Likert response	21
Chapter 3: Professional development workshops do make a difference...	
Fig. 1. Relationship between RTOP and Interactivity (IA) Scores	44
Fig. 2. Relationship between active/student centered approaches and RTOP/IA year since participation	45
Fig. 3. Relationship between active/student centered approaches and RTOP/IA year of participation	46
Fig. 4. Comparison of RTOP and Interactivity scores across the years	47
Fig. 5. Comparison of NASI and non-NASI by RTOP score	48
Fig. 6. Distribution of RTOP scores for NASI and control across categories	48
Fig. 7. Frequency of time spent on teaching behaviors from COPUS	50
Fig. 8. Frequencies of time spent of student behaviors from COPUS	51
Chapter 4: Exposure to reformed science teaching and peer evaluation may improve Future Faculty teaching	
Fig. 1. Types of Training GTAs have been exposed to	69
Fig. 2. Perceived Teaching Strategies Used During Instruction	69
Fig. 3. Elements Associated with Student-centeredness	70
Fig. 4. Future Faculty relationship between Pre RTOP and Interactivity Scores	71
Fig. 5. Future Faculty relationship between Post RTOP and Interactivity Scores	71
Fig. 6. Theme analysis of Teaching Feedback Guide question two	73
Fig. 7. Theme analysis of Teaching Feedback Guide question three	74

LIST OF TABLES

Tables	Page
Chapter 2: Participation in an adaption of the National Academies Summer Institute...	
Table 1. Sample agenda for NASI@WVU	11
Table 2. Value of the NASI@WVU sessions	15
Table 3. Participants agree with SI-related teaching values after participating in the NASI@WVU	17
Table 4. Anticipated challenges and supports to implementing Scientific Teaching practices	19
Chapter 4: Exposure to reformed science teaching and peer evaluation may improve Future Faculty teaching	
Table 1. GTA Cohort with RTOP and timing analysis pre and post scores	72

LIST OF VALIDATED SURVEYS

Surveys	
Chapter 2: Participation in an adaption of the National Academies Summer Institute...	
	National Academies Summer Institute – Exit Survey
	National Academies Summer Institute – 1YR Follow-up Survey
Chapter 3: Professional development workshops do make a difference...	
	Reformed Teaching Observation Protocol (Sawada & Piburn, 2002)
	Classroom Observation Protocol for Undergraduate STEM (Smith et al., 2004)
Chapter 4: Exposure to reformed science teaching and peer evaluation may improve Future Faculty teaching	
	Teaching Feedback Guide adapted from J. Momsen (FIRST IV)

Chapter 1: Introduction

Transforming Science Education

The US was known once for its scientific and engineering innovations, but this has been on the decline since the latter part of the twentieth century (Atkinson and Mayo, 2010). In 1983, the National Research Council (NRC) published the landmark report, *A Nation at Risk: The Imperative for Educational Reform*, describing the state of education in the United States at all levels: primary, secondary and postsecondary. The report examined the quality of both teaching and learning occurring in the classroom. At that time, American students scored last in 7 out of 19 tests taken by the United States and other countries. Thirty-five years later, we are struggling with the same conundrum of a prosperous nation, such as the US, ranking 41st in the world in math and 25th in the sciences behind a number of developing nations according to the Programme for International Student Assessment (PISA, <https://nces.ed.gov/surveys/pisa/>) carried out by the Organisation for Economic Co-operation and Development (OECD, 2015).

Because our global rankings in math and science have been in decline, the need for workers who have expertise in these areas is increasing. The need for science, technology, engineering and math (STEM) majors is outpacing their production with a projected need for one million trained STEM professionals by this year (2018) according to a report by the President's Council of Advisors on Science and Technology (PCAST, 2012). Several longitudinal studies have found that students are interested in STEM majors but tend to switch to other majors primarily in the first two years of college in large part due to continued use of passive lecture techniques (Chen, 2013; Graham, *et al.*, 2013; Handelsman, 2011; Handelsman *et al.*, 2004; Lowell, *et al.*, 2009; Mervis, 2010; Seymour and Hewitt, 1997; Stains *et al.*, 2018; Tobias, 1990; Wood, 2009). They are not challenged to analyze, apply or evaluate scenarios that can be applied to real life situations which leaves them without the

skills needed to apply their knowledge (AAAS, 2011; Burrowes, 2003, DiCarlo, 2006; Lujan and DiCarlo, 2005, NRC, 2003). For many non-scientists, their first formal encounter with the process or nature of science is in an introductory science class at the undergraduate level. A focus on the “what” and not the “why” or “how” can turn them off to the sciences, although some may have been genuinely interested in the field (Seymour and Hewitt, 1997; Tobias, 1990; Wood, 2009).

Reforming science education in order to effectively train all students has become a national imperative (AAAS, 2011; NRC, 2003; Wood, 2009; Woodin *et al.*, 2009). Innovations and reform in teaching the nature of science can help engage students in ways that will better prepare majors for STEM careers and will empower non-majors to make future science-related decisions (Handelsman, 2011). In 2000, the NRC published the report “*How Students Learn*” that compiled evidence from educational research to inform recommendations on teaching practices. Learning requires students to gain a body of “usable knowledge” that can be accessed to solve problems rather than to accumulate a list of disconnected facts (NRC, 2000). Evidence suggests that using active learning to engage students’ prior knowledge to answer questions and solve problems is the most effective approach to help students transition from novice to expert (Freeman *et al.*, 2014; NRC, 2000).

Use of evidence-based teaching strategies, such as active, cooperative, and inquiry-based learning, leads to improvements in student learning gains, critical thinking skills and confidence (Beichner, 2003; Freeman *et al.*, 2014; Knight and Wood, 2005, Schroeder *et al.*, 2007, Timmerman *et al.*, 2008). Improved student learning through the use of reformed methodologies is not limited to the classroom. Replacing traditional laboratories with inquiry-based and/or authentic research experiences allows students to develop an appreciation for the nature of science and to develop necessary science process skills

(Brownell *et al.*, 2015; Gormally *et al.*, 2011; Olimpo *et al.*, 2016; Wallace *et al.*, 2003).

Given the effectiveness of these strategies, why aren't more schools rushing to implement them? Amidst other factors such as institutional cultures and disciplinary identities that prioritize research over teaching (Brownell and Tanner, 2012), most post-secondary instructors lack the necessary training. Post-secondary education is the only level of the educational system where formal pedagogical training is rarely a required part of graduate training. K-12 teachers undergo considerable pedagogical training and are required to be evaluated and certified before they are allowed to teach. While the salaries and future funding of K-12 teachers can be affected by a failure to produce the appropriate percentage of passing students, post-secondary faculty are rarely held to such a standard (Tanner and Allen, 2006).

STEM graduate programs tend to provide rigorous training in research, but typically have no requirement for pedagogical training, leaving faculty to revert to the methods used on them as students (Romm *et al.*, 2010; Stains, 2018). Without formal training, faculty can easily fall prey to the misconception that if they succeeded in traditional classrooms their students should as well. They tend to blame failure on the students rather than the method of teaching. One educator related her epiphany when students wrote in evaluations that they slept through the lectures all semester and they would prefer not to do that (Burrowes, 2003). Statements like these are strewn throughout articles on science education reform. Students are becoming both more aware that poor teaching poses a challenge to success in STEM and more vocal about classes that focus on memorization over all else (Sirum and Madigan, 2010).

Effective pedagogical training for current and future faculty (we refer to graduate students and post-docs collectively here as future faculty, whose intention is to become faculty even if it is not likely all will) is needed to create and sustain the reform necessary to improve post-secondary STEM education. The call for reform in biology education made in

Bio2010: Transforming Undergraduate Education for Future Research Scientists (NRC , 2003) helped improve awareness of the need for pedagogical training and led to the development of the National Academies Summer Institutes (NASI) founded in 2004. NASI was designed to address the lack of formal pedagogical training in STEM graduate programs by training science faculty in the best evidence-based teaching methods through an approach termed scientific teaching. Scientific teaching distills best practices from the learning sciences and combines them with an approach that incorporates the nature and process of science into the classroom (Handelsman *et al.*, 2006).

Since about 80% of students pursuing graduate degrees hope to enter faculty positions (Tanner and Allen, 2006), formal training in evidence-based teaching approaches that mimic graduate training in research - investigation, analysis and evaluation - should be standard practice (Handelsman *et al.*, 2007). While many reform advocates call for ‘standards of teaching practice’ at the post-secondary level, none have yet been made mandatory (Tanner and Allen, 2006) and few institutions have formal pedagogical training programs for graduates or their graduate teaching assistants (GTAs) (Romm *et al.*, 2010).

Focus of Study

The focus of this dissertation is the impact of pedagogical interventions on teaching approaches employed by current and future STEM faculty. Specifically, I am interested in how expanding the scope of the NASI from biology to all STEM fields and from current faculty at research institutions to current and future faculty from any post-secondary institution will affect its impact. I hypothesize that the format and approach are significantly robust to improve teaching knowledge, attitudes and behaviors of participants from a variety of STEM disciplines and ranks. In addition, I hypothesize that incorporating elements of the SI into a training program for graduate teaching assistants (TAs) for an introductory biology teaching laboratory will improve their use of active, student-centered teaching.

To this end, chapters 2-4 of this dissertation evaluate A) the impact of the first full-length regional offshoot of the SI, the NASI@WVU, on teaching knowledge, skills, attitudes and behaviors of its participants (Ch. 2, 3) and B) the impact of an adapted SI workshop and peer-mentoring on teaching behaviors of graduate TAs teaching in the introductory biology program (BIOL 115/117) in the Biology Department of West Virginia University (Ch. 4).

Chapter 2: Participation in an adaption of the National Academies Summer Institute on Undergraduate Biology Education to serve current and future STEM faculty leads to changes in teaching skills and behaviors.

ABSTRACT

In the absence of pedagogical training, faculty tend to rely more heavily on traditional, passive teaching methods which underprepare students for careers in STEM fields. In response to a call from the National Research Council, the National Academies Summer Institute (SI) was created to provide formal training in evidence-based practices for Biology educators from research universities. Participants valued the interactive format of the Summer Institute and reported changes in their teaching knowledge, skills and behaviors. In order to expand the impact of the Summer Institutes, a regional spin-off was adapted to serve current *and* future *STEM* faculty and teaching staff from *any* post-secondary institution. The regional Summer Institute replicated the general interactive format of the original SI which requires participants to apply the new methods they are learning by developing and peer-reviewing teaching units during the course of the weeklong institute. A new interdisciplinary workshop was developed for the regional SI and larger and/or interdisciplinary teams were recruited from 2- and 4-year post-secondary institutions. Over the four years that this regional SI was offered, participants reported significant increases in teaching knowledge, skills and implementation of active learning strategies. Some areas of increased knowledge, skills and implementation included active learning, measuring student learning and using learning outcomes to guide teaching. One of the most well-received sessions of the SI was scientific teaching. Ninety four percent of participants valued this session, while 75.9% believed that engagement in learning is a highly valued skill for successful active-learning/student-centered STEM classrooms. The findings from this study confirm that intensive immersive STEM workshops like SI are instrumental in motivating faculty to use scientific teaching in their classrooms.

INTRODUCTION

The need for science, technology, engineering and math (STEM) professionals is outpacing their production by the American higher education system (President's Council of Advisors on Science and Technology, 2012). This is not caused by a lack of interest in STEM by students. While nearly 30% of students seeking bachelor's degrees initially choose STEM fields, almost half of these leave STEM before graduation (Chen, 2013). Many students who begin college as STEM majors find themselves leaving, often in the first two years, because the teaching methods are less than favorable (Chen, 2013; Graham, *et al.*, 2013; Handelsman, 2011; Handelsman *et al.*, 2004; Lowell, *et al.*, 2009; Mervis, 2010; Seymour and Hewitt, 1997; Tobias, 1990; Wood, 2009). Among other things, students encounter a focus on memorization and regurgitation of content that turns them off to the sciences. Those who stay are not well-equipped by this training to pursue STEM careers (DiCarlo, 2006; Lujan and DiCarlo, 2005).

Post-secondary science education is in need of pedagogical reform that will engage students and improve learning across the board (American Associate for the Advancement of Science, 2011; National Research Council, 2003; President's Council of Advisors on Science and Technology, 2012; Wood, 2009; Woodin *et al.*, 2009). When educators use evidence from research to inform course design and implement teaching strategies such as active, cooperative, inquiry-based learning, students show improvements in learning gains, critical thinking skills and confidence (Beichner, 2003; Freeman *et al.*, 2014; Knight and Wood, 2005, Schroeder *et al.*, 2007, Timmerman *et al.*, 2008). In the absence of formal pedagogical training, college STEM instructors often use the same techniques they observed as students (Romm *et al.*, 2010; Stains *et al.*, 2018). Many believe that if they were able to succeed in a traditional lecture classroom then their students should be able to do the same. Even students are aware of the need for improvements in teacher training (Schroeder *et al.*, 2007). In a

2010 survey, students cited poor teaching techniques as a primary reason for leaving STEM disciplines (Sirum and Madigan, 2010). They are becoming increasingly unwilling to tolerate classrooms where they have to memorize large bodies of information that they cannot recall in the following semester (Sirum and Madigan, 2010).

While there is overwhelming evidence that active, inquiry-based learning in undergraduate biology education improves learning (Freeman *et al.*, 2014), college science faculty and teaching staff tend to lack formal pedagogical training to successfully implement reformed approaches. Faculty still tend to rely on traditional teaching methods in undergraduate biology classrooms and laboratories (Stains *et al.*, 2018). *Bio2010: Transforming Undergraduate Education for Future Research Scientists* (National Research Council, 2003) spawned significant change in the awareness of and training for pedagogical reform in undergraduate biology education through the National Academies Summer Institutes (SI) founded in 2004. The SIs are 5-day immersion-style workshops that train STEM faculty in evidence-based teaching methods through an approach termed scientific teaching.

The centrally located SIs were held at the University of Wisconsin, Madison and trained 30-60 biology educators per year between 2004-2011. Participants found the immersive program to be valuable in helping them gain awareness, knowledge and skills in evidence-based teaching practices in addition to reporting changes in their teaching practices once they returned to their home institutions (Pfund *et al.*, 2009). A 2.5 day mini-institute model of the SIs, Scientific Teaching, Assessment, and Resources (STAR), was piloted at Louisiana State University from 2007-2011. This pilot extended the SI reach to biology faculty from diverse post-secondary schools. Like the central SI, participants, regardless of institution type, reported that the immersive approach increased their use of evidence-based practices and improved student learning (Gregg *et al.*, 2013).

In order to further extend the impact of the SIs, the first full-length regional SI was piloted at West Virginia University (WVU) between 2010 and 2014. This regional institute replicated the National Academies Summer Institute on Undergraduate Biology Education model of an intensive 5-day experience consisting of groups of participants working together to create instructional materials using the evidence-based teaching methods introduced during interactive workshops. The National Academies Summer Institute at West Virginia University (NASI @WVU) expanded the original model in three notable ways. While the original workshop focused on biology faculty from primarily research institutions, NASI@WVU trained both current and future faculty from STEM departments at any post-secondary institution. Four consecutive regional SIs provided training to current and future STEM faculty in scientific teaching. Due to the success of this regional pilot, the SIs transitioned from a central to a fully regional model in 2012, with regional SIs being offered at 4-7 sites around the country each year since.

The regional expansion dramatically increased the number of participants served by the SIs. Here, we report on the impact of the NASI@WVU on participants from a variety of ranks, disciplines and institution types. When developing this regional offshoot, we hypothesized that the immersive, interactive format of the original SI was sufficiently robust to have a positive impact on participant attitudes, knowledge and reported changes in classroom activity, regardless of discipline or rank, when implemented by the second generation trainers who were participants of the original SI.

METHODS

Description of the NASI@WVU

NASI@WVU was a week-long, intensive professional development program where participants learned evidence-based teaching strategies to improve student learning. During morning sessions, participants engaged in interactive workshops to gain knowledge and

practice in reformed teaching strategies and backward design (Wiggins and McTighe, 1998, Table 1). During the afternoon sessions, they worked in groups to create teaching modules using their newly-acquired skills and strategies from the morning sessions. As a regional offshoot of the original SI, NASI@WVU provided nationally-recognized speakers for the workshops and experienced summer institute alumni to facilitate the group work. During the week, groups peer-reviewed one another's teaching modules. These preliminary presentations provided formative feedback which allowed groups to gauge their progress and revise their teaching materials prior to the final presentation. On the last day of the institute, groups presented their teachable materials to the entire institute. The final presentations served as summative assessments both to foster buy-in by participants and to assess the level of proficiency participants had achieved in scientific teaching over the course of the week. Participants agreed to use one or more teaching modules and teaching approaches at their home institutions during the following year. They also agreed to send one representative to a follow-up meeting the following year to report on successes and challenges.

Participants in the NASI@WVU were recruited through the dissemination of an email invitation with program information to the Deans and Chairs of STEM departments at post-secondary institutions in West Virginia and surrounding states, and to alumni of the SIs. Applicants completed both individual and team applications modeled after those of the original SI. Institutional teams of current and future faculty were given preference over individuals, however, single participants were accepted in special circumstances. Institutional teams could be single or multidisciplinary and, unlike the original SI, there was no cap on team size.

Table 1. Sample agenda for NASI@WVU

Time/Day	Monday	Tuesday	Wednesday	Thursday	Friday
8:30	Scientific Teaching	Formative Assessment	Integrating Across Disciplines	Assessing Student-centered Classrooms	Presentation Preparation
10:15	Backward Design & Summative Assessment	Active Learning	Diversity and Unconscious Bias	Assessment to Research Dissemination	Group Presentations
11:15					
1:00	Group Work	Group Work	Group Work	Group Work Group pictures	Reflection, Wrap-up and Evaluation
4:30			Group Share	Group Share	

Evaluation

In order to compare the performance of the NASI@WVU with the original SI, we used surveys adapted from those used to evaluate the original SI with the help of S. Miller and C. Pribenow (U. Wisconsin). These metrics were used to explore the impact of the NASI@WVU on teaching knowledge, attitudes and behaviors. Both surveys were administered online using Survey Monkey (www.surveymonkey.com).

The first metric, referred to as the Exit Survey, was modeled after the instrument used by the original SI, to capture participants' immediate perceptions of the value of the SI and their changes in teaching knowledge and attitudes (Pfund et al., 2009). The Exit Survey was administered to participants within a month of completion of the SI. It consisted of five sections: participant demographics, perceived value of the elements of the SI, perceived impact of the SI on knowledge, attitudes and intended teaching behavior, and expected challenges in the coming year. The questions were primarily multiple choice and Likert-scale with a small number of free response questions.

The second metric, the 1-Year Follow-Up survey, was used to collect information about the strategies that the participants implemented during the year following the SI and the

successes and challenges faced during implementation. Participants completed this survey one year after participation in NASI@WVU. This survey consisted of five sections: participant demographic, the impact of the pedagogy interventions on teaching since leaving the institute, how the institute has affected teaching (before and after) and if there had been any dissemination of pedagogy information the year after the institute

Responses to open-ended prompts in the two surveys were subjected to thematic analysis using a grounded theory approach (Braun and Clarke, 2006; Regassa and Morrison-Shetlar, 2009). Raters reviewed the responses to develop response codes. Codes and emergent themes were then discussed and refined. Responses were then binned into similar themes and a frequency for each theme or sub-theme was calculated.

Data Analysis

For the Exit Survey and the 1-Year Follow-up Survey, frequency and descriptive statistics were calculated for responses to Likert-scale questions either by averaging the numerical score for each question or by calculating the average percentage response rate for each answer. The Kruskal-Wallis H test was used to compare independent populations across years. Mann-Whitney U-test was used to make pairwise comparisons between independent samples while the Wilcoxon signed rank test was used to make pairwise comparisons between dependent samples, such as pre-/post-responses. Responses by populations differing by discipline or rank were compared using Pearson's Chi-Squared. All statistical analysis was performed using the Statistical Package for Social Sciences (SPSS, version 25) software.

The NASI@WVU protocol was approved by the WVU Institutional Review Board (WVU IRB# H-21531).

RESULTS

Participant Demographics

The first full-length offshoot of the original National Academies Summer Institutes (SI, University of Wisconsin, Madison) was held on the campus of West Virginia University (WVU) in 2008. The format followed closely that of the original SI with an intensive, 5-day workshop combining active learning workshops and group work to create teaching modules using reformed pedagogical practices. Thirty-two participants consisting of graduate students, post-doctoral fellows, faculty and teaching staff attended from the WVU Department of Biology along with a small cohort of discipline-based education researchers from the other STEM departments within the WVU Eberly College of Arts and Sciences. With the help of a subset of that first discipline-based education research (DBER) cohort, the National Academies Summer Institute at WVU (NASI@WVU) was adapted for a STEM audience. In 2009, this newly adapted STEM Summer Institute was attended by thirty-one participants from STEM disciplines across the WVU campus. Funding was sought and acquired from the National Science Foundation to evaluate the impact of four consecutive regional Summer Institutes from 2010 to 2013 after the positive response by participants of the first two cohorts.

During those four summer institutes, an average of 29 participants per year (total=116; 59% female, 41% male) attended the regional SI from a variety of STEM disciplines and post-secondary institutions. Participants came predominantly from regional colleges and universities, although occasionally institutions from states outside SI's designated offshoot regions (e.g., California, Michigan, Florida) would send participant teams based on recommendations from alumni of the original SI. Participants hailed more often from doctoral institutions (62%) than from masters/baccalaureate institutions (33%) with a few participants from publishing companies (which serve undergraduate STEM

courses, 5%), but no representation from community colleges or associate-degree granting colleges. Biology and biology-related disciplines were the mostly highly represented STEM discipline, followed by Chemistry, Physics, Math and Engineering (Figure 1). The “Life Sciences Related” category includes participants from health science disciplines, health science education and communication in the health sciences. In addition to institutional and disciplinary diversity, participants varied by rank, years of teaching experience and training. Faculty (72%) made up the largest segment of the participant population followed by graduate students (24%) and post-doctoral fellows (4%) with a wide spread in years of teaching experience (<6 years = 36%, 6-10 years = 25%, 11-20 years = 30% and >20 years = 9%). For training, participants were more likely to have observed others teaching (87.3%) or attended professional development workshops (46.6%) than to have taken undergraduate (19%) or graduate (33.2%) pedagogy courses.

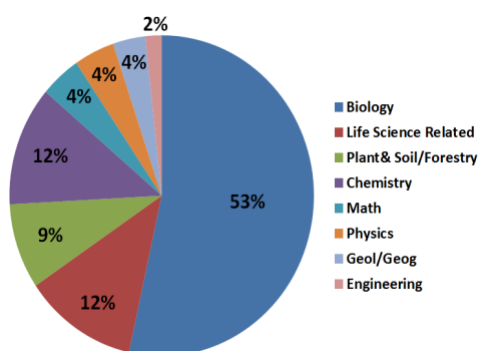


Figure 1. Representation of NASI@WVU attendees by discipline, n=116 for four consecutive years of the NASI@WVU, 2010-2013.

Perceived value and impact of participating in the regional National Academies Summer Institute.

Completion of an Exit Survey was an important aspect of gauging participants’ perceived benefits from the week-long summer institutes. Since the NASI@WVU was expanding its audience to include different ranks and different disciplines, in addition to being run by alumni of the original SI, we were interested to find out if NASI@WVU participants still valued the institute highly, as participants had for the original SI. Average

response rates for the Exit Survey were 82.8% with no individual year's rate falling below 65%. The majority of participants perceived the NASI workshops as valuable (Table 2) and would recommend the SI to a colleague (86%). There was no significant difference in perceived value of the SI sessions by rank (current versus future faculty; $p > 0.05$, Pearson's Chi-Squared test) or discipline (Biology versus other STEM, $p > 0.05$, Pearson's Chi-Squared test). The most highly-valued workshops addressed the introduction to scientific teaching and two of its three main themes (active learning and assessment). Similar to the original SI, the third theme of scientific teaching, diversity, was among the lower rated workshops, as was the Institutional Transformation workshop, during which participants made plans to disseminate what they learned at the SI on their home campuses.

Table 2. Perceived value of the NASI@WVU sessions by participants

SI Session	“Valuable”/“Extremely Valuable”
Scientific Teaching	92.8 ± 1.1
Assessment	94.0 ± 4.1
Active Learning	91.6 ± 3.5
How People Learn	88.4 ± 6.1
Diversity/Unconscious Bias	71.4 ± 9.1
Institutional Transformation	62.9 ± 13

Average percentage of respondents from Exit Surveys, 2010-2013, who found the sessions “valuable”/“extremely valuable”. Mean ± s.e.

For nearly a third of the respondents this was their first pedagogical workshop. The majority of those who had attended other workshops rated the SI positively in comparison (76%). Below are representative comments in response to the prompt “How does the SI compare to other professional development workshops you have attended?”

- *“It's the only workshop I've attended, but it's extraordinarily helpful compared to just being thrown into a classroom”*
- *“Far more extensive (and helpful) than any other workshop I've attended.”*
- *“This was far better, and more applicable, as it directly addressed methods and practical questions.”*
- *“This is the first science education workshop I've attended. I found it transformative; although I feel I still have a lot to learn I feel it has resulted in fundamental changes*

- *to the way I think about teaching.”*
- *“Similar but better than the other two I have attended because it focuses on science and had a practical focus.”*

Participants not only valued the SI workshops but reported that the experience improved their teaching knowledge and skills. The majority of respondents reported feeling confident that they could implement the teaching strategies they learned at the SI (70%). There was no significant difference in confidence levels across ranks or disciplines ($p > 0.05$, Pearson’s Chi-Squared test). When asked to compare their knowledge and skills before and after the SI, respondents reported significant increases in their ability to perform specific teaching practices taught at the SI (Figure 2, $p < 0.001$, Wilcoxon signed rank test). The practices that saw the biggest perceived improvements were creating teaching materials that: “elicit prior knowledge”, “include formative assessment”, “uncover student misconceptions” and “help students gauge their own learning”.

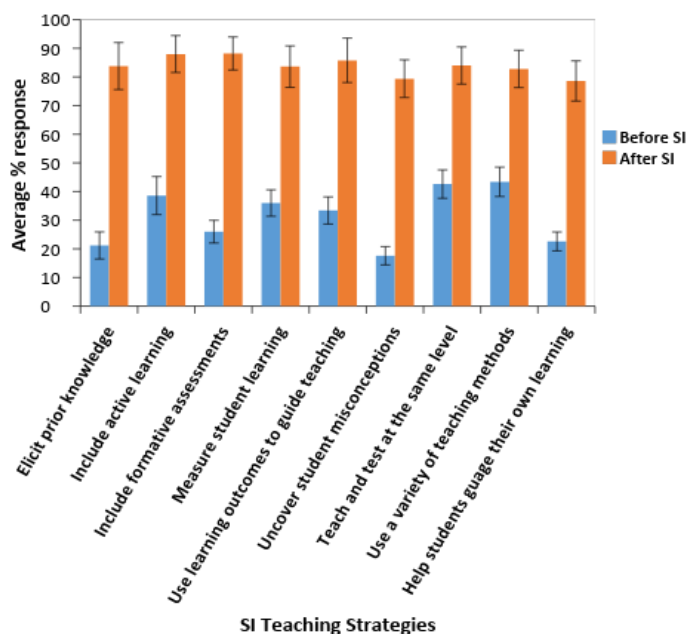


Figure 2. Perception of teaching skills differs before and after participation in the NASI@WVU. Response frequencies (Exit Survey) for each cohort/year (2010-13) who perceived themselves as high/very high for specified teaching skills, were averaged. Mean \pm s.e. Wilcoxon ranked sign test, $p < 0.001$.

In addition to teaching skills, respondents reported on their beliefs about teaching following participation in the Summer Institute. A higher percentage of respondents “agreed”/”strongly agreed” with statements that reflected the values of the SI, such as the importance of student engagement and using evidence to determine student learning, than statements that ran counter to SI values like valuing content coverage above all else (Table 3).

Table 3. Participants agree with SI-related teaching values after participating in the NASI@WVU.

Teaching Value Statement	Agree/ Strongly Agree
Covering CONTENT is the most important part of my class	25.8 ±3.6
Students should be ENGAGED in learning	75.9 ±4.2
Students should learn what I INTEND	65.5 ±4.1
I should have EVIDENCE that my students are learning what I intend	63.8 ±4.4
Students should know the GOALS in my course	74.2 ±4.1

Percentage of respondents from the Exit Survey who “Agree”/”Strongly Agree” with a given teaching value statement. Mean ± s.e.

While respondents felt confident in the knowledge and skills gained at the SI, they were able to speculate on potential challenges and sources of support as they strove to implement the new teaching strategies. Current and future faculty identified an overlapping set of challenges and barriers (Table 4.) that share commonalities with those identified by participants of the original SI (Pfund, 2009). Following are representative statements in response to an open-ended prompt to identify potential challenges to and sources of support for the implementation and dissemination of SI practices.

Challenges:

- *It will take some time to adopt the skill and ideas of SI*
- *Team-taught courses are the biggest challenge. The worst-case scenario is only one out of five faculty in a course uses active learning--how will that work?*
- *As a grad student, some of the courses are managed by faculty who don't want to radically change their courses to include more active learning and better assessment*
- *Busy schedules, old habits, possible student resistance to some of the ideas.*

- *Not a 'large' group of educators in our department, faculty may be resistant to any professional development.*

Supports:

- *Having a teachable tidbit that I can directly use in my class is a big help.*
- *I now feel confident in using "Clickers" in my classroom and have had time to learn about and reflect on other methods of engaging students.*
- *Networking with other teachers who have attended the summer institute (outside of my college) and communicating with those from my college that also attended to get ideas/suggestions.*
- *My department, which is supportive of these efforts, and our students, who have already been appreciative of the changes we've made.*
- *It will help to monitor the course and students understanding*

The majority of items that were identified as a challenge by one respondent were identified as a support by another. For example, money, teaching technology and student buy-in were about as likely to be considered a support when participants had them as a challenge when they did not. Only five of the fourteen sub-categories of items were viewed as only a challenge or a support. The two groups placed different emphases on the importance of certain items as either supports or challenges. Time was the most common challenge identified by current faculty followed by institutional support, while opportunities to teach and personal expertise/skills were more commonly identified by future faculty. The two groups did agree on the importance of peer support as the most commonly identified support and lack thereof as a challenge. Additionally, future faculty were more likely to identify access to teaching materials, and course supports as supports than current faculty who focused on institutional support as second only to peer support.

Table 4. Anticipated challenges and supports to implementing Scientific Teaching practices. CF = current faculty, FF = future faculty.

Item Category		Support		Challenge	
		CF	FF	CF	FF
Resources	Time	5	--	29	10
	Money	--	5	3	--
	Teaching materials	7	20	3	--
	Teaching technology	5	--	9	--
Environment	Opportunities to teach	2	5	3	40
	Having control over one's course[s]	5	5	--	--
	Course Content	--	--	6	--
	Institutional support	17	5	17	--
	Peer support	39	40	14	15
	Student Buy-in	5	--	6	5
	Course support/Evaluation	2	10	--	--
Personal	Expertise/skill development	10	10	6	20
	Comfort implementing	--	--	3	--
	Comfort disseminating	--	--	3	10

Percentage of responses from the Exit Surveys that identified a specific challenge or support in response to an open-ended prompt. Supports, n=41 for current faculty (CF), n=20 for future faculty (FF); Challenges, n= 35 for CF, n=20 for FF.

Implementation in the year following the NASI@WVU

One year after attending the SI, all participants were asked to report on their teaching activities during the intervening year. Representatives from each institutional team also returned to WVU for a follow-up meeting to formally present their team's successes and challenges. The average response rate on the 1-Year Follow-up Survey was 55% with no individual year's rate falling below 30%. In the year following the SI, participants spent an average of 47% of their time teaching with faculty and instructional staff averaging 53%, and post-docs and graduate students averaging 27%. They spent the majority of that time teaching undergraduate courses for majors (67%) roughly equally distributed across introductory (66%) and upper level (68%) courses, while non-majors (34%) and graduate courses (28%) took up less of the time. Faculty spent an average of 25% of their time on research, while

post-docs and graduate students averaged 59%.

Many more current (68%) than future (16%) faculty reported being able to implement some aspect of Scientific Teaching (backward design, active learning, assessment, diversity/inclusivity) following the Summer Institute. When asked to compare their use of these Scientific Teaching strategies before and after attending the SI, participants indicated a significant increase in implementation of backward design, active learning diversity practices and integrating concepts from other disciplines (Figure 3A. Wilcoxon signed rank test, $p < 0.05$). In addition, participants were asked to give a more concrete accounting of their implementation of specific learning activities and formative assessments strategies such as whole class discussion, problem-solving, clicker questions, group work and learning objectives. Participants indicated their use of these strategies on the following scale: never, once a semester, a few times a semester, monthly, weekly, daily (Figure 3B). Current faculty implemented whole class discussion, group work and problem-solving more frequently than future faculty (Figure 3B. Kruskal-Wallis test, $p < 0.05$).

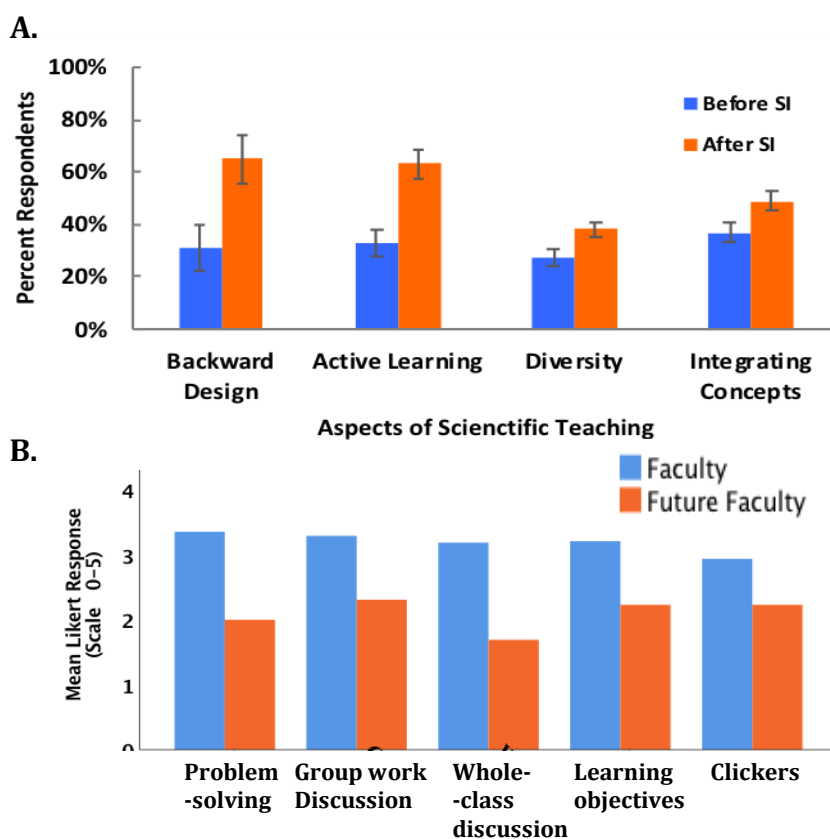


Figure 3. Reported implementation of Scientific Teaching strategies. **A)** Average percentage of respondents from the 1-Year Follow-up Surveys, 2010-13, who reported implementing aspects of scientific teaching in their teaching before and after the SI. Mean \pm s.e., Wilcoxon signed rank test, $p < 0.05$. **B)** Average Likert response to a prompt on the 1-Year Follow-Up survey asking participants to indicate frequency of implementation of specific learning activities and formative assessments. Likert scale: 0 = never, 1 = once/semester, 2 = few times/semester, 3 = monthly, 4 = weekly, 5 = daily. Blue bars = current faculty; red bars = future faculty. Kruskal-Wallis test, $p < 0.05$ for whole-class discussion, group work, and problem-solving.

In addition to responding to Likert scale questions about implementation of specific scientific teaching strategies, respondents were asked to reflect on the impact the SI had on their teaching in the year after the SI. Following are representative statements from that open-ended prompt.

- *I gained a sufficient foundational understanding of scientific teaching/active learning that I could begin to incorporate it into my teaching and introduce other faculty to it.*
- *I now know how to approach designing courses with effective learning strategies, and I can more effectively assess where students are having trouble learning, and I can address those trouble areas with effective tools.*
- *I am much more aware about assessment, and diversity. Also, I have a much better understanding about the concept and techniques related to scientific teaching and the student-centered learning approach.*

- *WVUSI is one of several workshops I have attended or facilitated where backwards design is a major theme. What was different about SI was the intensive time to focus on a teachable unit. I didn't end up teaching that unit b/c of about 6 missed snow day classes and a shift in content but I don't feel it was a waste for me b/c I used the same skill set in assessment design and alignment in some other activities that I did implement.*

Four common themes emerged from these responses: knowledge/skill acquisition, awareness of the importance of the SI, facilitating student improvement, and ability to integrate concepts from other disciplines. Responses from both current and future faculty were represented in each theme, however at different frequencies. Current faculty responses more often fell under the themes of awareness of the importance of the SI and integration of concepts from other disciplines (33% for both) followed by knowledge/skill acquisition (22%) and student improvement (11%), while future faculty showed almost the reverse trend with knowledge/skill acquisition and student awareness trending higher (30% each) than awareness of the importance of the SI and integration across disciplines (20% each).

Evaluating Student Learning and Affect

Using evidence-based teaching practices is a hallmark of scientific teaching and by extension the training provided by the SIs. A scholarly approach to teaching, *i.e.* gathering evidence to demonstrate student learning and affective changes in response to a course, is valued and promoted by the SI curriculum. As such, participants were asked to share the methods that they employed to measure student learning and/or effective response to their class during the year following the SI. The vast majority of participants (89%) used a combination of the institution-mandated student evaluation of instruction with another metric. Over a third reported using their own metric, a quarter used pre-/post-testing with validated concept inventories and six percent used a validated perception survey (Seymour et al., 2000). Only eleven percent of respondents reported using the institution-mandated student evaluation form as their sole form of course evaluation.

Respondents were also asked to estimate changes in nine areas of student attitudes and behaviors in response to their implementation of Scientific Teaching in their classrooms. For eight of the nine areas, nearly 50% or more of respondents reported that their students were “more”/“much more”: appreciative of the value of active learning (69%), willing to ask or answer questions (68%), enthusiastic about learning (62%), able to recognize when/what they did not understand (57%), appreciative of the relevance of course topics (54%), prepared for class (49%), likely to perform better on exams (48%). The majority felt that their students were about the same for attendance (63%) and their ability to take responsibility for their learning (57%).

Dissemination of Scientific Teaching

To expand the use of scientific teaching, participants were encouraged to disseminate the information learned at the summer institute on their home campuses. Forty-three percent of respondents, primarily current faculty, made an effort to disseminate scientific teaching to their colleagues. Of those, 36% of current faculty and 7% of future faculty were able to offer a Scientific Teaching event, including giving a seminar or workshop, on their own or another campus or professional meeting. Other dissemination efforts took the form of mentoring a colleague (40%), meeting regularly with colleagues to discuss SI techniques (32%), submitting an education grant proposal (12%) or science education manuscript (3%). A little over a quarter chose to take part in another form of professional development for teaching.

DISCUSSION

The overall picture represented by the data shows a clear positive relationship between participation in the NASI@WVU and the perception of improvement in attitude, knowledge and skills that can help participants move from traditional lecture to active, student-centered classes/classrooms. Scientific Teaching, the curriculum of the SIs, is an

evidence-based approach that provides a condensation of best practices from the education literature packaged for scientists who are already well-versed in empirical approaches to research. Even though the approach was developed to train biologists, there is nothing that should make it seem any less valuable or beneficial for current and future faculty from any STEM discipline. This is reflected in the responses of our participants who perceived the experience as highly valuable and beneficial to their teaching regardless of discipline or rank.

Another indication of the value of NASI@WVU was the increase in participation over the years. Between its inception in 2008 and its final year in 2013, participation increased by 200%. The majority of participants were current faculty (64%). Due to the push for improvements in science teaching at the undergraduate level, many universities are now placing more emphasis on quality teaching. New or potential faculty are finding that quality teaching is becoming more valued by their institutions (Dancy and Henderson, 2010) and reform in science classroom practices has become a requirement at some institutions. The need for training in evidence-based practices motivates both current and future faculty to take part in professional development in order to effectively implement reformed teaching in their classrooms. The efficacy of professional development paradigms has become the subject of many new research studies (*e.g.*, Stowell *et al.*, 2015).

Unlike the original SI, NASI@WVU served current and future STEM faculty from a variety of institution types. The number of future faculty (graduate students and post-doctoral fellows) attending the NASI@WVU increased so that future faculty made up 30% of all participants by the end. However, even with the similar recruitment approaches to all disciplines, Biology and biology-related disciplines still predominated. The most effective recruitment tool for the SIs seems to be word-of-mouth testimonials which would likely mean that the Biology focused origin would take longer to overcome. There was an increase in representation from other disciplines over the years.

While self-report surveys can provide convenient tools to measure teaching attitudes and practices, there are questions about their reliability, specifically that respondents may over-report changes in response to an intervention (Dancy, *et al.*, 2016; Ebert-May *et al.*, 2011; Kane, 2002). However, other studies have found that asking for more concrete estimations or having better alignment between questions and practices can lead to higher accuracy (Smith *et al.*, 2014; Walter *et al.*, 2016; Weigel *et al.*, 1974). As such, we feel confident in our participants' responses that indicate significant increases in the use of active learning techniques such as whole-class discussion, clicker questions or problem-solving on concrete scales of daily, weekly and monthly. While estimating how active or student-centered their class may be, could be difficult, reporting that you used a technique for the first time, or increased from once a month to once a class should be reliable determinations to make. Participants also indicated that certain workshops, Diversity and Institutional Transformation, were not as valuable and we saw this reflected in lower reported gains in confidence and implementation for those approaches.

One possible reason for the lower value reported for the institutional change workshop is that we are simply asking too much of participants to both implement newly acquired teaching skills and strategies in their courses and disseminate those approaches to their colleagues. Closely following the theme that time to implement was essential, lack of peer and institutional support were the biggest challenges identified by current faculty. In the wake of studies that have demonstrated lower than desired changes in adoption of evidence-based practices after professional development (Borrego and Henderson, 2014; Ebert-May, *et al.*, 2011; Henderson *et al.*, 2011) some have espoused the view that professional development does not work. However, studies show that professional development can lead to increases in teaching knowledge and skills along with motivation to change (Henderson *et al.*, 2011). Perhaps expecting professional development to also make up for institutional

barriers to change is unreasonable. The more effective approach will likely be to keep the professional development programs but pair them with elements to overcome institutional barriers to adoption of evidence-based practices. This approach is being used by new initiatives such as the Mobile Summer Institutes that use Henderson's four categories of change strategies model to add environmental elements to the pedagogical training already offered by the Sis (Henderson *et al.*, 2011). Given the differences in challenges and supports identified by the current and future faculty, initiatives that begin to address these issues should consider customizing interventions to the specific challenges faced by their audiences.

Limitations

During evaluation of NASI@WVU data it was recognized that an entrance survey on some of the items in the exit survey may have shed more light on the responses given by participants at the SI. Items that included reflecting on their current practices such as implementing aspects of scientific teaching in their classrooms and percentage of time they use teaching strategies may have garnered responses that were more true to their actual classroom behaviors and practices than post workshop.

Acknowledgements: We thank Sarah Miller and Christine Pribenow (University of Wisconsin) for help with adapting the original SI surveys for use by the NASI@WVU; Dr. Mark Graham (Yale University) for help with collection of survey responses in year 2012 and 2013; Drs. David Miller, Kasi Jackson and Michelle Richards-Babb (West Virginia University) for help in adapting the SI format for a STEM audience; Drs. Randall Phillis (University of Massachusetts, Amherst), Nancy DiIulio (Case Western Reserve University) and Michael Hanna (Rensselaer Polytechnic Institute) for help leading and facilitating the NASI@WVU workshops; Dr. Keith Garbutt (Oklahoma State University) for use of Honors College facilities to run the workshops, and Drs. Jonathan Cumming (WVU, then Chair of the Department of Biology) and Mary Ellen Mazey (Bowling Green State University, then

Dean of ECAS, WVU) for support and funding to pilot the 2008 and 2009 summer institutes at WVU; the National Science Foundation, Course Curriculum and Laboratory Improvement (CCLI) Phase II Award #0919800 PI, Withers.

REFERENCES

- American Associate for the Advancement of Science (2011) “*Vision and Change in Undergraduate Biology Education: A Call to Action.*” Washington, DC.
<http://visionandchange.org/> final report (accessed 9 January 2014).
- Beichner, R., & Saul, J. (2003). *Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) project.* Proceedings of the International School of Physics.
- Borrego, M., & Henderson, C. (2014). Increasing the Use of Evidence-Based Teaching in STEM Higher Education: A Comparison of Eight Change Strategies. *Journal of Engineering Education*, 103: 220–252.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.
- Chen, X. (2013). “*STEM Attrition: College Students’ Paths Into and Out of STEM Fields (NCES 2014-001)*”. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78, 105.
- Dancy, M., Henderson, C., & Turpen, C. (2016) How faculty learn about and implement research-based instructional strategies: The case of Peer Instruction. *Physical Review Physics Education Research*, 12, 010110.
- DiCarlo, S. E. (2006). Cell biology should be taught as science is practised. *National Reviews Molecular Cell Biology*, 7, 290–296.
- Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt H., Wenderoth, M-P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the*

United States of America, 111, 8410.

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). "Increased course structure improves performance in introductory biology." *CBE - Life Sciences Education*, 10 (2): 175-86.

Graham, M., Frederick, F., Byars-Winston, A., Hunter A-B., & Handelsman, J. (2013). Increasing Persistence of College Students in STEM. *Science*, 341:1455-1456.

Gregg, C., Ales, J., Pomarico, S., Wischusen, E., & Siebenaller, J. (2013). Scientific teaching targeting faculty from diverse institutions. *CBE - Life Sciences Education*, 12:383-93.

Handelsman, J., D. Ebert-May, R., Beichner, P., Bruns, A., Chang, R., DeHaan, J., ...Wood, W. (2004). Scientific Teaching. *Science*, 304: 521-522.

Handelsman, J. (2011). *The Revolution in Science Education: From Fringe Activity to National*. Center for the Integration of Research, Teaching and Learning. Madison, WI.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, 48:952-984.

Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, 72: 177-228.

Knight, J., & Wood, W. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4:298-310.

Lowell, B. L., Salzman, H., Bernstein, H., & Henderson, E. (2009). *Steady as She Goes? Three Generations of Students through The Science and Engineering Pipeline*. Paper presented at the Annual Meetings of the Association for Public Policy Analysis and Management, Washington, DC.

Lujan, H. L., DiCarlo, S. E. (2005). Too much teaching, not enough learning: what is the solution? *Advances in Physiological Education*, 30, 17–22.

Mervis, J. (2010). Better Intro Courses Seen as Key to Reducing Attrition of STEM Majors. *Science*, 330:306.

National Research Council. (2003) *Bio2010: Transforming undergraduate education for future research biologists*. Washington, DC: Nation Academy press.

Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D.,...
Handelsman, J. (2009). Summer institute to improve university science teaching. *Science*, 324: 470.

President's Council of Advisors on Science and Technology. (2012). "*Engage to Excel: Producing one million additional college graduates with degrees in science, technology, engineering and mathematics.*" Washington, DC: U.S. Government Office of Science and Technology.

Regassa, B., Morrison-Shetlar, A. I. (2009). Student Learning in a Project-Based Molecular Biology Course. *Journal of College Science Teaching*, 38:58-67

Romm, I., Gordon-Messer, S., & Kosinski-Collins, M. (2010). Educating Young Educators: A Pedagogical Internship for Undergraduate Teaching Assistants. *CBE – Life Sciences Education*: 80-86.

Schroeder, C., Scott, T., Tolson, H., Huang, T.-Y., Lee, Y.-H. (2007). A meta-analysis of national research: effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44:1436-1460.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO.: Westview Press.

Seymour, E., Wiese, D., Hunter, A., Daffinrud, S. M. (2000). *Creating a Better Mousetrap: On-line Student Assessment of their Learning Gains*. Paper presentation at the

National Meeting of the American Chemical Society, San Francisco, CA.

Sirum, K. L., & Madigan, D. (2010). Assessing How Science Faculty Learning Communities Promote Scientific Teaching. *Biochemistry and Molecular Biology Education*, 38: 197-206.

Stains, M., Harshman, J., Barker, M., Chasteen, S., Cole, R., DeChenne-Peters, S., ... Young, A. (2018) Anatomy of STEM teaching in North American universities. *Science*, 359: 1468.

Stowell, S. M. L., Churchill, A. C., Hund, A. K., Kelsey, K. C., Redmond, M. D., Seiter, S. A., & Barger, N. N. (2015). Transforming graduate training in STEM education. *Bulletin of the Ecological Society of America*, 96: 317–323.

Timmerman, B., Strickland, D., & Carstensen, S. (2008). Curricular reform and inquiry teaching in biology: where are our efforts most fruitfully invested?. *Integrative and Comparative Biology*, 48:226–240.

Tobias, S. (1990). *They're not dumb, they're different: Stalking the second tier*. Research Corp, Tucson, AZ.

Weigel, R. H., Vernon, D. T. A., & Tognacci, L. N. (1974). Specificity of the attitude as a determinant of attitude-behavior congruence. *Journal of Personality and Social Psychology*, 30(6), 724-728.

Wiggins, G., McTighe, J. (1998) *Understanding by Design*. Alexandria, VA: Association for Supervision and Curricular Development.

Wood, W. (2009). Innovations in Teaching Undergraduate Biology and Why We Need Them. *The Annual Review of Cell and Developmental Biology*, 25:5.1-5.20.

Woodin, T., Smith, D., & Allen, D. (2009). Transforming undergraduate biology education for all students: an action plan for the twenty-first century. *CBE - Life Sciences Education*, 8, 271–273.

Chapter 3: Professional development workshops do make a difference: increased adoption of active, student-centered teaching approaches following immersive training in scientific teaching.

ABSTRACT

Self-report metrics are often used to measure changes in teaching following participation in professional development. While these instruments provide information and insight into changes in knowledge, attitude and behaviors, they are susceptible to the reliability pitfalls of self-reporting, for example over-estimating personal change or providing aspirational responses. Therefore, combining these metrics with evaluation of teaching behaviors using validated observation protocols can give a more accurate picture of the extent of change. To that end, teaching videos of participants of a regional offshoot of the National Academies Summer Institute at West Virginia University (NASI@WVU) were evaluated using three metrics: the Reformed Teaching Observation Protocol (RTOP), the Classroom Observation Protocol for Undergraduate STEM (COPUS) and a simple timing analysis procedure. RTOP indicates the level of student-centeredness, COPUS provides a picture of classroom engagement through relative frequencies of various activities by both the student and instructor, and timing analysis measures the amount of time spent in passive and active modes. When compared to STEM faculty who did not attend the NASI@WVU, NASI participants are both more interactive and more student-centered. NASI participants RTOP scores averaged a category three signaling that while not quite fully-student centered participants were definitely implementing strategies. This was further reflected by interactivity scores positively correlated with RTOP scores and COPUS codes for both instructor and students that indicated the use and reception of active learning aspects in the classroom.

INTRODUCTION

Reforming STEM education so that it provides more inclusive and effective training for all students has become an issue of national urgency (AAAS, 2011; Anderson *et al.*, 2011; NRC, 2003; PCAST, 2012; Wood, 2009; Woodin *et al.*, 2009). Even in the face of overwhelming evidence that active learning improves student performance and retention (Freeman *et al.*, 2014), post-secondary STEM classes still are taught predominantly using passive lecture (Stains *et al.*, 2018) which contributes to student migration away from STEM fields (Seymour and Hewitt, 1997; Tobias, 1990). Lack of knowledge of the education literature (Fox and Hackman, 2003), on pedagogical training and other institutional barriers (Brownell and Tanner, 2012; Henderson and Dancy, 2007) contributes to there being very little to no change.

Many prominent initiatives like the National Academies Summer Institutes (SI, Handelsman *et al.*, 2004; Pfund *et al.*, 2009) and the Faculty for Reformed Science Teaching IV (FIRST IV, Ebert-May *et al.*, 2015) have sprung forth over the last 15 years to provide training in evidence-based practices and help facilitate their broad dissemination.

The National Academies Summer Institutes (Pfund *et al.*, 2009) , a week-long, immersive training program in scientific teaching, began in 2004 on the campus of the University of Wisconsin, Madison. SI was a concrete response to *Bio2010: Transforming undergraduate education for future research biologists* (National Research Council, 2003). Scientific teaching is a constructivist approach that relies on evidence-based practices, active learning and assessment, in the context of backward design (Wiggins and McTighe, 1998) to create inclusive, student-centered classes. Teams of biology faculty from research universities around the country participated in the program where they took part in interactive workshops on reformed teaching strategies and created learning activities for their classes. The program was highly valued by participants who perceived changes in their motivation,

knowledge, skill and teaching behavior as a result (Pfund *et al.*, 2009).

Two regional spin-offs, the Scientific Teaching, Assessment and Resources (STAR) mini-institute at Louisiana State University (LSU) and the full-length National Academies Summer Institute at West Virginia University (NASI@WVU), extended the reach of the original SI by diversifying the target audience to include a variety of post-secondary institutional types (STAR and NASI@WVU) and current and future faculty from any Science Technology Engineering and Mathematics (STEM) discipline (NASI@WVU). In both cases, participants continued to value the experience highly and report changes to teaching knowledge, skills and behaviors regardless of rank, discipline or institutional type (Gregg *et al.*, 2015; Chapter 2).

All three versions of the Summer Institute employed self-report metrics to measure changes in participants' teaching. Self-report instruments are convenient to use for both the subjects and the researcher. However, some studies have found that once participants return to their home campuses after taking part in professional development, they do not implement their newly acquired approaches at desired levels and they overestimate their level of adoption (Dancy *et al.*, 2016; Ebert-May, 2011). Other studies have found that the accuracy of self-reporting can be influenced by the specificity of the request parameters or alignment between codes and behaviors (Smith *et al.*, 2014; Walter *et al.*, 2016; Weigel *et al.*, 1974).

Directly observing a classroom provides important, objective measures of teaching behaviors. To help capture those behaviors in a reliable fashion, several observation protocols have been developed since the turn of the century. Aligning the measures of the instrument with the goals of the professional development program is critical for accurately evaluating the impact of the program. The goal of the SIs is to train its participants in active, student-centered teaching approaches.

The Reformed Teaching Observation Protocol (RTOP, Piburn *et al.* 2000; Sawada *et*

al., 2002) was designed to measure the level of reform occurring in the classrooms of preservice science and mathematics elementary and secondary teachers (Piburn *et al.*, 2000). The instrument is a rubric which captures student and teacher behaviors and elements of the classroom community to give an indication of the level of student-centeredness of the class. It consists of five levels or subscales (*lesson design and implementation, content, procedural knowledge, classroom culture, student/teacher relationships*) that focus on the level and nature of interactions between the instructor and students and between the students themselves (Sawada *et al.*, 2002; Lund *et al.*, 2015).

Each observation protocol has its strengths and weaknesses. For example, having the ability to measure student-centeredness is very useful in evaluating teaching effectiveness, but the complexity and nuance of the approach is not easy to objectively measure. For that reason, RTOP requires significant training for reliable, consistent deployment. On the other hand, simple timing analysis provides a straightforward measure of the amount of class time spent in an active mode (student-teacher or student-student interactions) rather than a passive one (teacher disseminating information). This metric is easy to use and requires minimal instruction, but paints a fairly simple picture of the classroom that does not reveal the level of student-centeredness. Using active learning typically makes a class more student-centered because getting students engaged at any level reveals something about their learning. However, active learning can certainly be done in more or less student-centered fashions. For example, two classes could spend the same amount of time in an active mode, one using multiple clicker questions where the instructor gave the answer and rationale after each question and the second using fewer questions but requiring students to discuss and defend their answers. The second is clearly more student-centered. Student thinking is more clearly revealed and students are practicing desired skills like providing rationale for answers. Timing analysis alone would not capture this distinction.

The Classroom Observation Protocol for Undergraduate STEM (COPUS, Smith *et al.*, 2014), bridges this gap somewhat. Like timing analysis it measures the frequency of more concretely observable classroom behaviors which makes it relatively easy to use. Observers check boxes in a table every time they observe a given behavior during two-minute intervals. Unlike timing analysis, the frequencies it produces are relative. An instructor could lecture for thirty seconds in every two-minute interval and be deemed to lecture 100% of the time. However, it does give a richer picture than timing analysis and is easier to use consistently than RTOP (Lund *et al.*, 2015; Smith *et al.*, 2014). As such, we decided to use a combination of the three metrics to both get a more complete picture of our subjects and to determine how the outputs of the three metrics compare to one another.

In order to determine the impact of the NASI@WVU on the teaching behaviors of its participants, we evaluated the teaching videos of participants who successfully completed NASI@WVU and compared them to STEM faculty who had not taken part in the professional development program.

METHODS

Study Participants and NASI@WVU program

The participants for this study were current and future faculty from a variety of STEM disciplines and institutions who successfully completed the NASI@WVU during its six year tenure ($n = 176$, pilot SI = 2008-09, NSF-funded NASI@WVU = 2010-13). The NASI@WVU was adapted from the original SI held at the University of Wisconsin-Madison 2004-2011 (Handelsman *et al.*, 2004; Pfund *et al.*, 2009). NASI@WVU reproduced the immersive, interactive 5-day professional development program consisting of morning workshops on elements of Scientific Teaching – Diversity, Assessment and Active learning – in the context of Backward Design (Wiggins and McTighe, 1998) combined with afternoon

working groups where participants practiced creating teaching materials using their newly gained knowledge and skills. Participants gave formative and summative presentations of their teaching materials during the week for which they received feedback from their peers, allowing them to leave the workshop with new knowledge and skills in addition to peer-reviewed teaching materials. Participants typically took part in the NASI@WVU as part of an institutional team, although a few participants were the only representatives from their institution at least for that year. Institutional teams sent a representative to a follow-up meeting one year after the NASI to report on successes and challenges. NASI@WVU differed from the original SI not in format but in audience. It expanded the reach of the SIs to STEM disciplines other than biology, to post-secondary institutions other than research universities and to future faculty in addition to current faculty. (“Future faculty” in this paper refers to graduate students and postdoctoral fellows. While we recognize that many of these individuals will not, in fact, become faculty, a large majority of them had that intention at the time they participated in the NASI@WVU and therefore provide a valuable perspective.)

Evaluation

In order to evaluate the impact of the NASI@WVU, participants responded to two surveys (Chapter 2) and a subset (n=38) agreed to submit a video of their teaching for analysis using three evaluation metrics: Reformed Teaching Observation Protocol (RTOP, Piburn *et al.*, 2000, Sawada *et al.*, 2002), the Classroom Observation Protocol for Undergraduate STEM (COPUS, Smith *et al.*, 2013) and timing analysis (personal communication R. Phillis, University of Massachusetts, Amherst), a simple type of discourse analysis (Hjelm, 2013) that measures the amount of class spent in active modes. This study focuses on the evaluation of teaching behaviors using these validated metrics.

In the year following completion of the NASI@WVU, each new cohort received email requests to complete surveys or submit teaching videos as part of the evaluation

program. Submitting a video for analysis was voluntary and not required as part of participation in the NASI@WVU. The first two pilot years of the NASI@WVU served current and future faculty from WVU only, while roughly one-third to one-half of each of the four cohorts of participants who attended the NSF-funded NASI@WVU between 2010 and 2013 were from WVU. In addition to the sizeable proportion of participants from WVU, the local participants could have their classes recorded by the research team making the process much more convenient. As a result, the majority of teaching videos were submitted by participants from WVU. Each yearly request was met with limited success. However, around the end of the program, between 2012 and 2014, requests for teaching videos were sent to all cohorts and the positive response rate increased. As such, teaching videos represent a variety of times between successfully completing the NASI@WVU and the recording, allowing us to investigate the effect of time since participation on implementation. For comparison, a random selection of faculty from STEM departments at WVU, who had not participated in the SI, were asked to submit teaching videos. A subset of those agreed to participate and allow their class to be recorded ($n = 9$).

To ease anxiety about being recorded, the instructor could choose the class being recorded. For WVU participants, a videorecorder was placed in the rear of the room by a member of the research team, at an angle that allowed for the observer to see the front of the classroom and a segment of the student population from the rear. Instructors were provided with a script that they could use to explain the situation to the class and allow any students to move out of the field of view of the camera if they so desired. This was provided to those instructors who set up their own video recording equipment and were not directly videotaped by the research team. The member of the research team would leave after successfully setting up the equipment and return after class to retrieve the equipment. The principle investigator safe-guarded the teaching videos and made them available to trained raters as needed. Videos

were assigned a random number and submitted to trained raters without any additional identifying information other than what could be seen in the video itself. Raters, therefore, may be privy to the identity of the instructor (although not all participants were from WVU), but were not alerted as to whether, and if so how recently, the instructor had participated in the NASI@WVU.

Participants from institutions other than WVU were required to set up their own videotaping equipment to videotape a class of their choice. The videos were sent to the research team either on CD via postal service or uploaded by the participant to a secure internet hard drive where they were retrieved by the research team for labeling and storage.

Rater Training and Video Evaluation

The teaching videos were evaluated using RTOP (Sawada *et al.*, 2002), COPUS (Smith *et al.*, 2013) and timing analysis. Five faculty from STEM disciplines (Biology, Chemistry, Mathematics) and Women's and Gender Studies at WVU and two graduate students in the research program of the primary investigator were trained with the two observation protocols.

RTOP is a scoring rubric that measures reformed teaching practices with elements that focus on the teacher, the students, their interactions and the environment of the class (Sawada *et al.*, 2002). While RTOP values active learning, the practices it emphasizes fall more into the category of student-centered or learner-centered teaching. The scoring rubric consists of five areas: Lesson Design and Implementation, Propositional knowledge, Procedural Knowledge, Communicative Interactions and Student/Teacher Relationships. Each area consists of 5 items with a rating scale from zero to four giving the total rubric a range from 0-100 across the following five categories:

- I. 0-30, Lecture only;
- II. 31-45, Lecture with minor student engagement;

- III. 46-60, A broad range of significant student engagement;
- IV. 61-75 Student engagement in this category is distinguished by critique and experimentation;
- V. 76-100 This last category represents “unhindered open inquiry”.

All but the first category span fifteen points allowing for sub-categorization of high, medium and low in each category. The first two categories are indicative of teacher-centered classes while the last three represent increasing levels of student-centeredness.

For RTOP, raters went through three rounds of training, first with Drs. Diane Ebert-May and Terry Derting, the primary investigators of the NSF-funded Faculty Institutes for Reforming Science Teaching IV (FIRST IV) program (Ebert-May *et al.*, 2011; 2015). FIRST IV was an NSF-funded professional development program for post-doctoral fellows in STEM whose training program was adapted from the SIs. As part of the FIRST IV evaluation, Drs. Ebert-May and Derting used RTOP and trained multiple RTOP raters, two of which (one faculty and one graduate student) were part of the NASI@WVU team of raters. Due to delays in procuring the participant teaching videos, the NASI team of raters were retrained twice a year later by the graduate researcher who had served as a rater for FIRST IV using the same process as Derting and Ebert-May. All rounds of training consisted of using RTOP to rate segments of videos comparing scores until the team reached an acceptable level of intraclass correlation coefficient (ICC) ($ICC \geq 0.70$; Gwet, 2010). All seven raters attended at least two of three training sessions. Once the acceptable level of consistency was established, trainers were randomly assigned teaching videos to rate. As the raters were rating their videos asynchronously, the first three and last three videos (based on random number assignment) were assigned to all raters to check whether ratings were consistent from beginning to end of the rating period. The reliability measure for the final three videos showed that although ratings were done asynchronously across different time scales, the rater reliability was still good ($ICC_{[2,7]} = 0.989$). Each teaching video was randomly assigned two raters and if scores

by those raters were more than one sub-category apart, a third rater was randomly assigned to resolve the discrepancy. Reliability was determined using an intraclass correlation as an estimator of interrater reliability ($ICC_{[2,7]} = 0.986$). ICC is comparable to Sawada *et al.* (2002) ICC of 0.97 done on the entire RTOP instrument during the validation process for the RTOP rubric.

COPUS is an observation protocol where the rater checks a list of student and teacher behaviors at two-minute intervals to get a relative frequency of the specific behaviors. Behaviors in the protocol range from more passive/teacher centered activities such as teacher talking or student listening to more engaging activities such as instructor posing a clicker question or students engaging in group work. While the relative frequencies produced by this are not as accurate as timing analysis, *i.e.* an instructor could lecture for thirty seconds of every two-minute period and get a 100% lecture output, it does provide a much richer picture of the diversity of activities employed in a classroom.

The seven raters were also trained with COPUS according to the instructions supplied through the supplemental materials (Smith *et al.*, 2014). COPUS training occurred concurrently with the second round of RTOP training. Briefly, all raters coded a training video together at specified intervals of increasing length in order to discuss their choices and both become more comfortable with the rubric and achieve an acceptable rate of inter-rater reliability. Raters were then asked to independently rate two videos referred to as training videos and raters were not aware these were part of the video population for this study. ICC were calculated on Cronbach's alpha score, $ICC_{[2,5]} = 0.708$ (student codes), $ICC_{[2,5]} = 0.705$ (instructor codes) (0.5 to 0.75 considered moderate reliability) (Koo and Li, 2016). At the half-way mark of coding, raters were again asked to code three additional videos. In the same manner, raters were unaware that these two were a part of the actual video population of the study. On this round of IRR is $ICC_{[2,6]} = 0.816$ (student codes; 0.75 to 0.90 considered good

reliability) and ICC $_{[2,6]} = 0.915$ (instructor codes; considered excellent reliability).

Raters used all three rubrics to score each video, allowing for comparisons across rubrics. Since timing analysis gave a more accurate picture of the amount of time spent in an active or passive mode, Cohen's kappa was calculated only for active COPUS codes for both teacher and student. Rater pairings for this study were random. In other words, a random selection process was used to determine rater 1 and rater 2 for each video, so identical pairs of raters may have rated some of their subset of all videos but did not rate each video within the subset of videos assigned to them. As such a two-way random effects model, with multiple k raters/measurements and consistent measures ICC(2, k) (Koo and Li, 2016) was utilized for a Cronbach's alpha (95% confidence intervals).

The timing analysis procedure employed for this study is a simple form of discourse analysis (Hjelm, 2013). Briefly, student-to-student or student-instructor interactions were considered active, while dissemination of information by the instructor was considered passive. The most subjective determination made with this method is deciding when an answer given by the instructor, in response to a student question, transitions into lecture mode. Raters were instructed to label dissemination of information that continued past the point needed to answer the specific question as lecture and thus a transition into passive teaching. For timing analysis, observers used a simple table to record the time point at which the teaching mode changed from a passive, e.g. lecture or demonstration, to an active mode, e.g., group work, clicker questions or whole class discussion. The percentage of time the class spent in an active mode was determined by dividing the active time by the total time of the class session and multiplying by one hundred.

Data/Statistical Analysis

Data were collected from teaching videos using three instruments; RTOP, COPUS and timing analysis, as described above. For each measure, descriptive statistics were carried

out to calculate the mean and standard error when appropriate. Given the bounded nature of RTOP (0-100 scale), COPUS and timing analysis (percentages) scores, non-parametric statistical tests were used to do pairwise comparisons between independent samples (Mann-Whitney U test), analysis of variance across multiple samples (Kruskal-Wallis H test) and correlations between measures (Spearman rank correlation). All data analyses were carried out using the Statistical Package for Social Sciences (SPSS, version 25).

The NASI@WVU protocol was approved by the WVU Institutional Review Board (WVU IRB# H-21531).

RESULTS

Measuring teaching behavior

Between 2008 and 2013, 179 participants took part in the two pilot and four NSF-funded years of the first full-length regional offshoot of the National Academies Summer Institute on the campus of West Virginia University. In order to determine the impact that the NASI@WVU had on teaching behaviors, we collected teaching videos from participants who successfully completed the program and evaluated them using three observation metrics: the Reformed Teaching Observation protocol (Sawada and Piburn, 2002), the Classroom Observation Protocol for Undergraduate STEM (COPUS, Smith *et al.*, 2014) and timing analysis, a simple version of discourse analysis (Hjelm, 2013).

Thirty-eight participants (21% of the total NASI@WVU population) submitted videotapes for evaluation. The vast majority were from WVU (79%) with slightly more females (58%) than males (42%) and many more faculty members (76%) than graduate students (18%) or post-doctoral fellows (5%). Requests for teaching videos were sent to an equivalent number of randomly selected faculty members from the STEM departments at WVU who had not taken part in the NASI@WVU. Nine (Male=8, Female=1) of those

requests resulted in acquisition of teaching videos for evaluation. Given the different formats of teaching laboratories compared to lecture classes, we separated the two formats. Four of the thirty-eight teaching videos from NASI participants and two of the nine from the control group were teaching laboratories.

Since each teaching video was rated with multiple observation protocols, we were interested to know how the RTOP and timing analysis related to one another. There was a strong positive correlation between how interactive a class was (% active using timing analysis) and how student-centered it was (RTOP score) for both NASI participants and control teachers (Figure 1) (r_s represents r score for non-parametric Spearman correlation test) $r_s=0.918$, $p<0.05$).

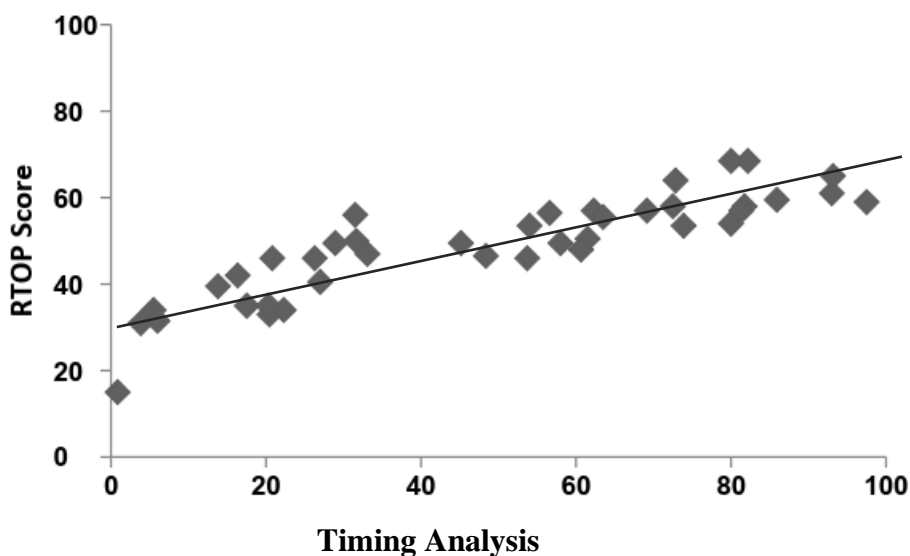


Figure 1. The relationship between RTOP and Interactivity Scores, where percentage of time in an active mode is calculated from timing analysis. Points are from all teaching videos (NASI and non-NASI) excluding those from teaching laboratories. Spearman rank test, $r_s=0.918$, $p<0.05$)

In addition to looking at the strength of relationship between the outputs of the metrics, we were curious about whether there was a difference in the level of adoption of active, student-centered strategies by participants as a function of time since participating in NASI. There was a moderate negative relationship between time since participation in NASI and RTOP score (Figure 2A, $r_s= 0.444$, $p<0.05$) and level of interactivity (Figure 2B, $r_s=$

0.468, $p < 0.05$). The highest RTOP scores occurred for participants that were within 1-2 years of participating in NASI while the lowest occurred for participants who participated in the SI five years prior to videotaping. The same trend is seen when relating level of interactivity to years since attending NASI.

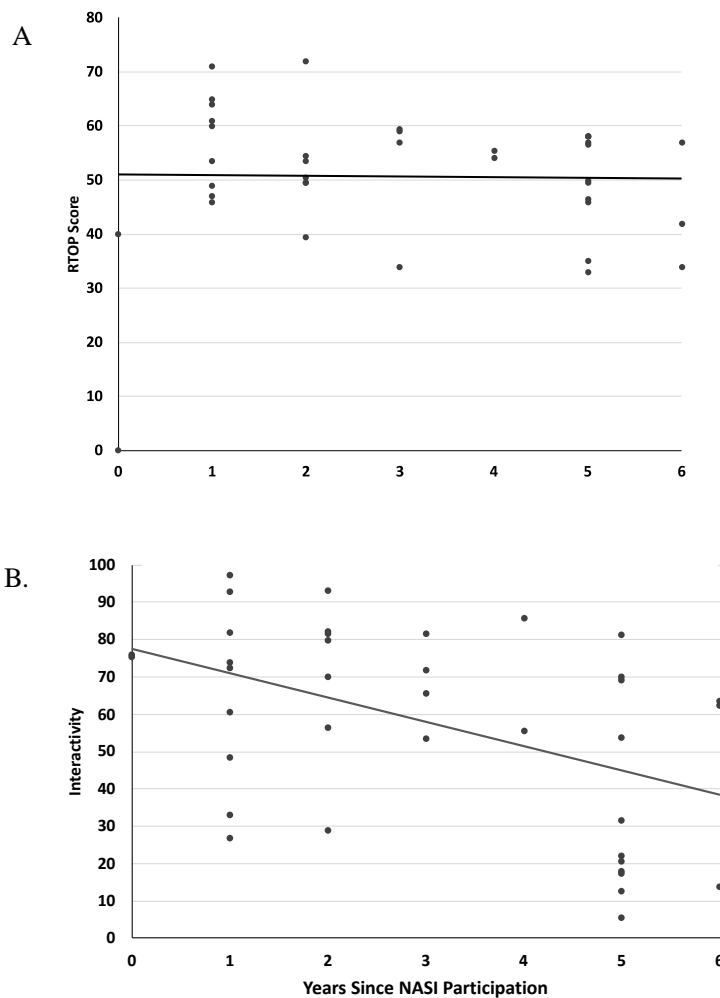


Figure 2. The relationship between active/student-centered approaches and years since participation in the SI. A) RTOP vs years since participation in the NASI@WVU, Spearman rank test, $r_s = 0.444$, $p < 0.05$; B) Level of interactivity vs. years since participation in NASI, Spearman rank test, $r_s = 0.468$, $p < 0.05$.

This trend reverses when comparing the year of participation with the two measures, RTOP (Figure 3A, $r_s = 0.450$, $p < 0.05$) and interactivity (Figure 3B, $r_s = 0.448$, $p < 0.05$). There is a moderate positive relationship between the year that participants attended NASI and their RTOP and Interactivity measures. More participants from NASI@WVU in 2013 have higher RTOP scores than any other year, while participants from 2008 and 2009, the two pilot years, had some of the lowest scores on these two metrics.

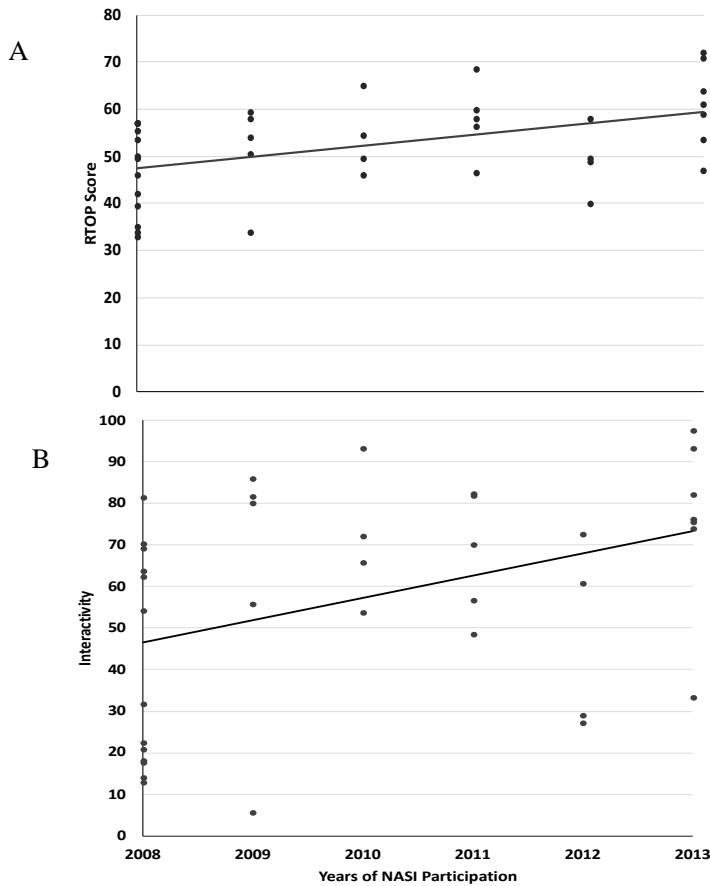


Figure 3. The relationship between active/student-centered approached and year of participation in the SI. A) RTOP vs the Year of participation in the NASI@WVU. Spearman rank test ($r_s= 0.450$, $p<0.05$). B) Level of interactivity vs. Year of participation in the NASI@WVU. Spearman rank test ($r_s= 0.448$, $p<0.05$).

When participants are separated based on their years since participating in the NASI or the year of their participation in the SI, there is a significant difference in the mean RTOP scores. There are no significant differences in the mean level of interactivity (Kruskal-Wallis, RTOP $p<0.05$, Interactivity $p>0.05$) over time. Participants from years 2008 and 2009, the pilot years, were also the participants who had the longest time since participating in the NASI. The lowest RTOP scores came from these two cohorts.

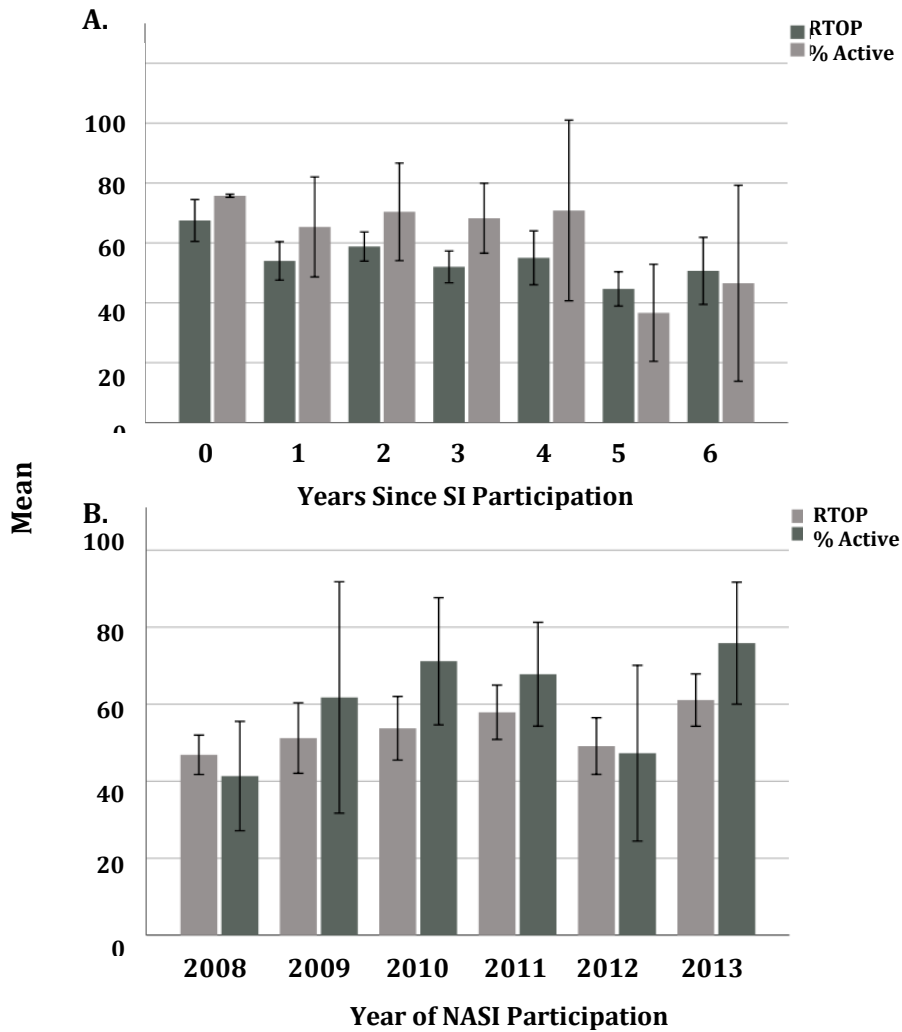


Figure 4. Comparison of RTOP and Interactivity scores across the years A) since participation in the SI (0 =2, 1 n= 7, 2 n=7, 3n=2, 4 n=2, 5 n=10, 6 n=4) and B) of participation in the SI (2008 n=13, 2009 n=4, 2010 n=3, 2011 n=5, 2012 n=4, 2013 n=5). Mean \pm s.e. Kruskal-Wallis H test, RTOP, $p < 0.051$; Interactivity, $p > 0.05$.

When comparing participants from the NASI@WVU with randomly selected STEM faculty from WVU who did not take part in the SI (Figure 5), there was a significant difference between the groups for both RTOP scores ($p=0.023$, $p < 0.05$) and level of interactivity in the classroom ($p=0.024$, $p < 0.05$). On average NASI participants scored 49 on the RTOP scale (category 3) and used 50% of class time for active learning. There was much more variation in the control group with two subjects exhibiting scores on RTOP and timing analysis similar to NASI participants.

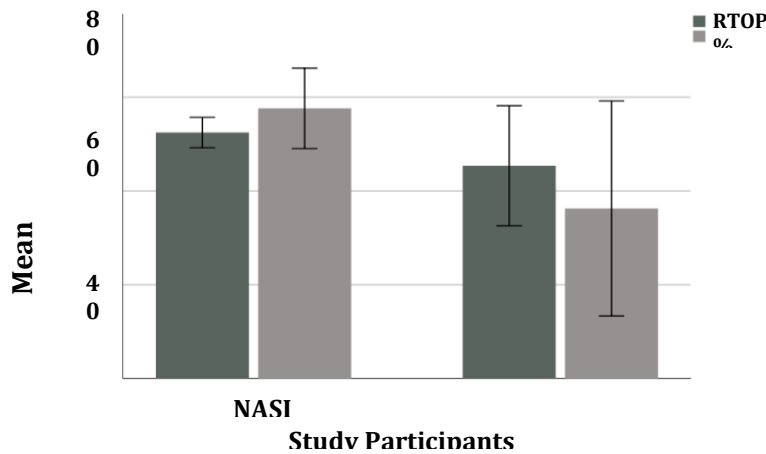


Figure 5. Comparison of NASI and non-NASI participants by RTOP score (dark gray bars) and timing analysis scores (gray bars). Bars represent the mean \pm s.e. Mann-Whitney U, $p < 0.05$.

RTOP not only provides a numerical score, but also separates teaching into distinct categories, Category 1 (15-30), Category 2 (31-45), Category 3 (46-60), Category 4 (61-75), Category 5 (>75). The lower categories are considered to be more teacher-centered while the upper categories are progressively more student-centered. The majority of RTOP scores for NASI participants (22/34) occurred in Category 3, with nearly equal numbers distributed across Categories 2 (7), 4 (5) and 5 (5) and none in Category 1 (Figure 6A). Categories 1 and 2 are not considered to represent teacher-centered classes. The control group had representatives in all Categories except 5, with the most in Category 2 (Figure 6B).

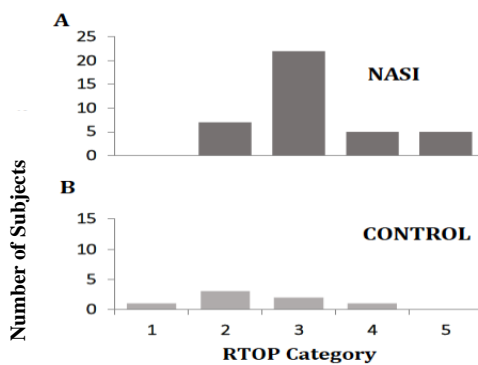


Figure 6. Distribution of RTOP scores for NASI participants and control educators across the five RTOP categories. Bars represent numbers of subjects A) NASI (dark gray bars) and B) control (light gray bars) educators who scored in the different RTOP categories.

Finally, teaching videos by NASI participants and non-NASI STEM teachers were also evaluated using COPUS. This metric can provide a richer indication of the specific forms that active learning takes in a classroom. Formatively, it can provide a rich profile for an individual to get a snapshot of their own classes. However, for means of comparing across groups of people, we only compared the frequencies of codes (specific activities from the observation protocol that represent engagement of students or active teaching by the instructor). Since we have more accurate measures of the amount of class time spent in active/passive modes from timing analysis, we ignored codes for these passive activities. Figure 7 shows the frequency profiles of NASI and control teachers for guiding active learning. The codes from COPUS that represent these teaching behaviors are: PQ – posing a non-clicker question, CQ – posing a clicker questions, AnQ – answering a student question, MG – moving around the classroom during group work, 1o1 – one-on-one interactions with groups during group work. NASI participants averaged more time in all codes except for answering questions when compared to non-NASI instructors (Figure 7), spending two to three times more class time posing questions and moving around the room interacting with groups during group work.

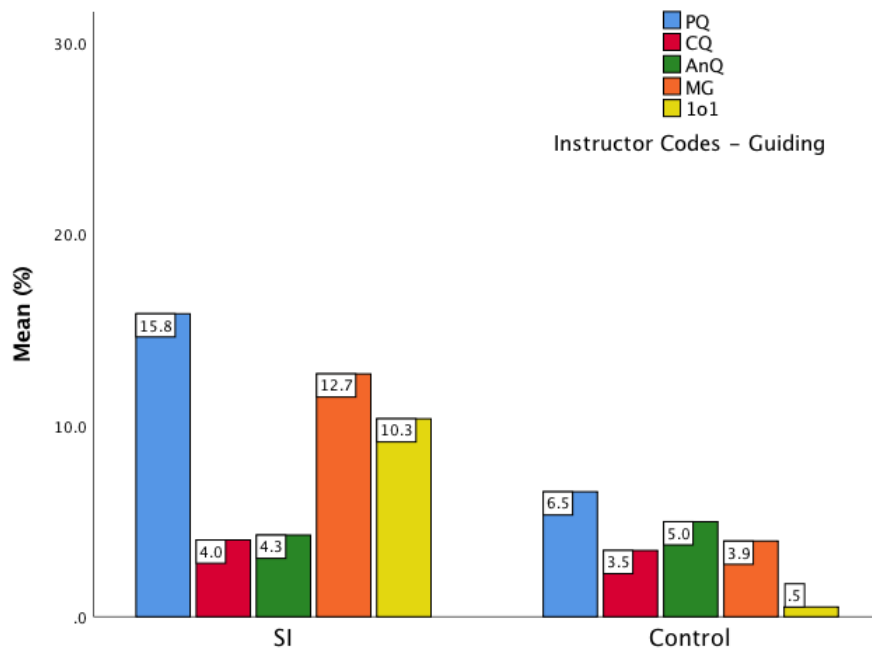


Figure 7. Time spent performing teaching behaviors that promote active learning. Bars represent mean percentage of time spent in specified teaching modes from COPUS by NASI (SI) and educators who did not participate in NASI (Control). PQ = posing a non-clicker question, CQ = posing a clicker question, AnQ = answering a question, MG = moving around the room during group work, 1o1 = working one-on-one with groups during learning activities.

We separated the active student codes into two categories, student working and student talking (Smith *et al.*, 2015). Student working includes individual thinking/problem-solving (IND), clicker question discussion in groups (CG), group work with a worksheet (WG), group activity not associated with clicker question or worksheet (OG), and making a prediction (Prd). NASI participants spent more time engaging students in individual thinking/problem solving, group-work with worksheets and even more group work time not related to worksheets or clickers, while non-NASI participants engaged students more often with clicker group work (Figure 8A). For codes related to students talking: AnQ – answering a question, SQ – asking a questions, WC – whole class discussion, and SP students presenting, the students of NASI participants answered questions over twice as often as the students of non-NASI educators while time spent in the other activities was relatively similar (Figure 8B).

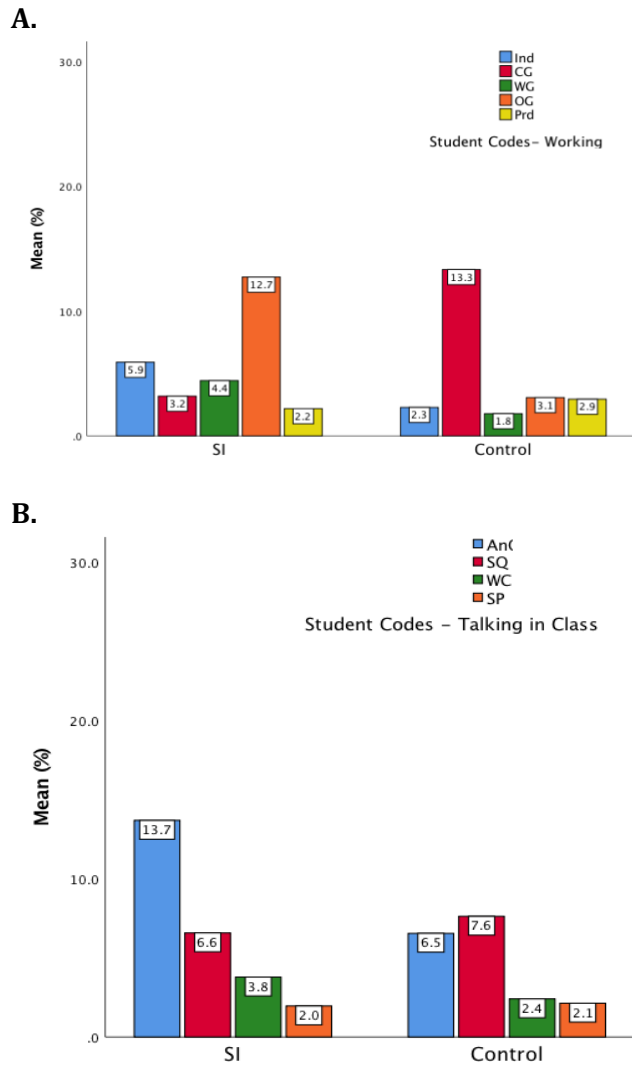


Figure 8. Student time spent A) engaged in learning activities, (Ind = individual thinking or problem solving, CG = group discussion or work on a clicker question, (WG = group work on a worksheet activity, OG = group not worksheet or clicker related, Prd = making a predication) and B) communicating with the class (AnQ = answering a question, SQ = asking a question, WC = whole-class discussion, SP = presenting to the class). Bars represent mean percentage of time spent in specified learning or communication activities from COPUS by students of NASI (SI) and of educators who did not participate in NASI (Control).

DISCUSSION

Professional development has gotten a less than desirable reputation lately. Because of the ease of use of self-report metrics, professional development endeavors often rely on these instruments to get an indication of changes in participants' attitudes, motivation, and teaching knowledge, skills and behaviors. In 2011, Ebert-May and others employed RTOP, a validated observation protocol, to evaluate teaching videos by participants from two national

professional development programs (Ebert-May *et al.*, 2011). The main finding was that instructors scored lower than expected on RTOP which was interpreted as over-reporting of their level of adoption of active learning by participants. However, one issue is that educators were estimating their levels of implementation of active learning and being evaluated with an observation protocol that focuses more on student-centered approaches. Certainly, use of active learning is part of the RTOP scores. Our data show that there is a strong relationship between levels of active learning in the class and RTOP scores. However, participants using equivalent levels of active learning, can have different levels of student-centeredness. Using RTOP as the only metric to determine the efficacy of professional development endeavors could cause researchers to dismiss programs that foster important early steps in the transition from passive, teacher-centered to active, student-centered instruction. In early stages of adoption, participants could be making important first steps in acquiring skills and comfort with active learning before they fully embrace the sophistication of a student-centered approach.

Self-report is more accurate when respondents are asked to report on more specific or concrete actions or behaviors (Weigel, 1972). Also, when there is closer alignment between the behavior being reported and the metric being used, over-reporting is less of an issue (Smith *et al.*, 2014). Our participants perceived that they increased their skill and use of concrete types of learning activities or teaching approaches (Chapter 2). Our findings with three different observation protocols, confirm this perception. The majority of our participants fell into RTOP category three, considered a student-centered category, and had an average RTOP of 49; both of which are higher than that found for subjects of the Ebert-May *et al.* 2011 study. Our participants spent, on average, nearly 60% of their class time in an active mode. This reinforces that the perceived changes in their teaching behavior and will lead to improving their students' performance and retention (Freeman *et al.*, 2014).

Additionally, our COPUS data show that participants are implementing a variety of learning activities that engage students. To truly evaluate the effectiveness of professional development programs requires metrics that capture implementation of evidence-based practices at a variety of levels and modes. Single metrics, especially ones like RTOP, that measure more nuanced, less concrete behaviors, might unintentionally miss important developmental steps in the transition to more student-centered classes.

The negative relationship between the time since participation in the SI and adoption of active, student-learning strategies is interesting. Earlier cohort participants (years 1 and 2) had lower RTOP scores than participants who had just been out of the SI for over a year or two. One might expect that more time to practice and/or develop activities would lead to higher adoption rates. However several factors could be playing a role. First, the early years (years 1 & 2) were pilot years of the program where the curriculum was being adapted to local workshop leaders and to a new STEM audience. Also, at that time, the newly hired WVU DBER faculty member in the Biology Department was just beginning to make changes to the departmental and institutional culture, so the campus environment was different than it was for later participants taking the NASI strategies to their classrooms.

The positive relationship between year at the NASI and increasing adoption may hint at a general increase in the awareness and knowledge by faculty of the need for teaching reform.

The last cohort of the SI may have been exposed to the reports from various science bodies, like the National Academies or the American Association for the Advancement of science that called for reform in college STEM education. It could also indicate that there is a need for continuous inoculation and/or support to continue to increase adoption over time (Owens *et al.*, 2018). This latter view is shared by many new programs and researchers who are bringing an understanding of organizational change to bear on the barriers to adoption of

evidence-based practices in higher education. Charles Henderson's four categories of change model emphasizes the need for a focus on both the individual and their surroundings (Borrego and Henderson, 2014). This approach is being adopted by the new Mobile Summer Institutes, an adaptation of the NASI based on Henderson's model.

Limitations

In chapter 4 we assess the merits scientific teaching for the purposes of a graduate teaching assistant training program. In this model we collected both pre and post video recordings of the graduate teaching assistants. It is possible that the design of this study could be improved if participants attending the SI workshops included a video of their teaching prior to exposure to NASI. We received video recordings from 21% of the total NASI participants with many participants alluding to the fact that they were not comfortable taping themselves teaching as yet. While having a pre video to compare to post scientific teaching workshop would change the scope of the results it may also cause a further decrease in participation of SI workshops. The goal is to have more people motivated to train and a stipulation of providing video recording prior to attendance may seem a burden to some participants.

How do we encourage continued use of evidence-based practices?

We can look at the improvement in adoption of active, student-centered pedagogies by STEM instructors as the beginning of a journey toward student-centered classes. Owens *et al.* (2018) conducted a 3 year study of a department-wide professional development program that began with all faculty members of the department taking part in a week long institute similar to the SI. They carried out observations of classrooms, documented the use of active learning techniques by participants and provided support to faculty as needed. This approach addresses the needs of the individual for training and iterative evaluation and reflection on teaching and provides support from the environment to help overcome institutional barriers

that may hinder adoption (Brownell and Tanner, 2012).

Acknowledgements: We thank Nicholas Wilbur (Utah Valley University) and Drs. Diane Ebert-May (Michigan State University), Terry Derting (Murray State University) for training in the use of RTOP; Nicholas Wilbur (Utah Valley University) and Drs. David Miller, Kasi Jackson and Michelle Richards-Babb, Katrina Stewart (West Virginia University) for help reviewing teaching videos; the National Science Foundation, Course Curriculum and Laboratory Improvement (CCLI) Phase II Award #0919800, PI Withers.

REFERENCES

- American Associate for the Advancement of Science (2011). “*Vision and Change in Undergraduate Biology Education: A Call to Action.*” Washington, DC.
- Anderson, W. A., Banerjee, U., Drennan, C. L., Elgin, S. C. R., Epstein, I. R., Handelsman J, ... Warner, I. M. (2011). Changing the culture of science education at research universities. *Science*, 331, 152–153.
- Borrego, M., & Henderson, C. (2014). Increasing the Use of Evidence-Based Teaching in STEM Higher Education: A Comparison of Eight Change Strategies. *Journal of Engineering Education*, 103: 220–252.
- Brownell, S. E., & Tanner, K. D. (2012). Barriers to Faculty Pedagogical Change: Lack of Training, Time, Incentives, and...Tensions with Professional Identity? *CBE - Life Sciences Education*, 11(4), 339–346.
- Dancy, M, Henderson, C., & Turpen, C., (2016). How faculty learn about and implement research-based instructional strategies: The case of Peer Instruction. *Physical Review Physics Education Research*, 12, 010110.
- Ebert-May, D., Derting, T., Henkel, T., Middlemis Maher, J., Momsen, J., Arnold, B., & Passmore, H. (2015). Breaking the cycle: Future faculty begin teaching with learner-centered strategies after professional development. *CBE - Life Sciences Education*, 14:1-12.
- Ebert-May, D., Derting, T., Hodder, J., Momsen, J., Long, T., & Jardeleza, S. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *BioScience*, 61:550-558.
- Fox, M. A., & Hackerman, N. (2003). *Evaluating and improving undergraduate teaching in science, technology ; engineering, and mathematics.* Washington, DC: National Academies Press.
- Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt H., &

Wenderoth, M-P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 8410.

Gregg, C., Ales, J. D., Pomarica, S., Wischusen, E., & Siebenaller, J. (2015). Scientific teaching targeting faculty from diverse institutions. *CBE - Life Sciences Education*, 12:383-393.

Gwet, K. L. (2010). *How to Compute Intraclass Correlation with MS EXCEL: A Practical Guide to Inter-Rater Reliability Assessment for Quantitative Data*. Gaithersburg, MD: Advanced Analytics.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... & Wood, W. (2004). Scientific Teaching. *Science* 304: 521-522.

Handelman, J. (2011). *The Revolution in Science Education: From Fringe Activity to National*. Center for the Integration of Research, Teaching and Learning. Madison, WI.

Henderson, C., & Dancy, M., (2007). Barriers to the use of research-based instructional strategies: the influence of both individual and situational characteristics. *Physical Review Physics Education Research*, 3, 020102.

Hjelm, T. (2013). Empowering discourse: discourse analysis as method and practice in the sociology classroom. *Journal of Teaching in Higher Education*. 18:871-82.

Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The Best of Both Worlds: Building on the COPUS and RTOP Observation Protocols to Easily and Reliably Measure Various Levels of Reformed Instructional Practice. *CBE - Life Sciences Education*, 14 (2) 18.

National Research Council. (2003). *Bio2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.

Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D., ...

Handelsman, J. (2009). Summer institute to improve university science teaching. *Science*, 324: 470.

Piburn, M. D., Sawada, D., Falconer, K., Turley, J., Benford, R., Boom, I., & Judson, E. (2000). *Reformed Teaching Observation Protocol (RTOP): Reference Manual*. Arizona Collaborative for Excellence in the Preparation of Teachers Technical Report. Tempe, AZ.

President's Council of Advisors on Science and Technology. (2012). "*Engage to Excel: Producing one million additional college graduates with degrees in science, technology, engineering and mathematics.*" Washington, DC.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., Bloom, I. & Judson, E. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, 102, 245–253.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview Press.

Smith, M., Jones, F., Gilbert, S., & Wieman, C. (2014). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE - Life Sciences Education*, 12:618-627.

Stains, M., Harshman, J., Barker, M., Chasteen, S., Cole, R., DeChenne-Peters, S., ... Young, A. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359: 1468

Tobias, S. (1990). *They're not dumb, they're different: Stalking the second tier*. Research Corp, Tucson, AZ.

Walter, E. M., Henderson, C. R., Beach, A. L., & Williams, C. T. (2016). Introducing the Postsecondary Instructional Practices Survey (PIPS): A Concise, Interdisciplinary, and Easy-to-Score Survey. *CBE - Life Sciences Education*, 15(4), ar53.

Weigel, R. H., Vernon, D. T. A., & Tognacci, L. N. (1974). Specificity of the attitude as a determinant of attitude-behavior congruence. *Journal of Personality and Social Psychology*, 30(6), 724-728.

Wiggins, G., & McTighe, J. (1998). *Understanding by Design*. Alexandria, VA: Association for Supervision and Curricular Development.

Williams, C. T., Walter, E. M., Henderson, C., & Beach, A. L. (2015). Describing undergraduate STEM teaching practices: a comparison of instructor self-report instruments. *International Journal of STEM Education*, 2: 18.

Wood, W. (2009). Innovations in Teaching Undergraduate Biology and Why We Need Them. *The Annual Review of Cell and Developmental Biology*, 25:5.1-5.20.

Woodin, T., Smith, D., & Allen, D. (2009). Transforming undergraduate biology education for all students: an action plan for the twenty-first century. *CBE - Life Sciences Education*, 8, 271-273.

Chapter 4: Exposure to reformed science teaching and peer evaluation may improve Future Faculty teaching in STEM classes
ABSTRACT

Training programs for Science Technology Engineering and Mathematics (STEM) graduate teaching assistants (GTAs) are not as prevalent as STEM faculty professional development. Discipline-based education research should focus on GTA Professional Development Programs (PDP) that take into consideration the whole graduate experience. GTA PDP mostly develop methods that are geared toward addressing an essential lack of teaching abilities and possible solutions. This study used elements of scientific teaching (SI) and a model of peer evaluation techniques from Faculty Institutes for Reforming Science Teaching (FIRST IV) to provide GTAs with tools and feedback to help them become more active and student-centered. Teaching behaviors of program participants were measured using the Reformed Teaching Observation Protocol (RTOP) before and after peer evaluation. There were no significant changes in RTOP scores before and after peer review. GTAs scored high RTOP category three before peer review and just as high after. We suspect there was no change because laboratory classes were already reformed to provide higher levels of active learning than a lecture class. However, feedback from peer evaluation suggests that GTAs view peer support as a necessary component of PDP to see any meaningful change in their roles as future faculty. Peer evaluation and support seem to be the key to helping GTAs enter academia not struggling to find their teaching style.

INTRODUCTION

Across colleges and universities, Graduate Teaching Assistants (GTAs) teach a majority of major and non-major courses (Kendall and Schussler, 2012). Many of these are the laboratory sections of various STEM courses (Kendall and Schussler, 2012). Most GTAs are entering these classes as first-time instructors with little to no teaching experience. Some universities and colleges offer a range of graduate teaching programs to address the needs of their students. These may include programs offered by teaching and learning centers or pedagogical training as part of orientation prior to the start of the semester. A lack of consistency in professional development programs to address the graduate student teaching skills development has been identified in educational research reform (Kendall and Schussler, 2012). One problem is that these programs (Blouin and Moss, 2015) do not succeed in fully preparing GTAs for an academic teaching career after graduate school.

Vision and Change (AAAS, 2011) brought more attention to the conversations that were occurring around biology education reform at the undergraduate level to help provide “best practices” (Ledbetter, 2012) for teaching biology to undergraduate students (Chen *et al.*, 2013) to make undergraduate students in the US globally competitive. The report cited a need for effective teaching at the undergraduate level and the necessity for faculty to be trained to implement scientific teaching strategies. Recommendations centered on providing authentic science experiences for undergraduate students (American Association for the Advancement of Science, 2011). The report did examine the need to train graduate and postdoctoral students to use scientific teaching (Ledbetter, 2012; Schussler *et al.*, 2015). However, there is still a lack of substantial training programs being put in place by higher education institutions.

While efforts at reform are most often focused on Biology faculty members and instructors, STEM education researchers do recognize the need to effect change at the GTA

level, including pre-service teachers (Ledbetter, 2012). There is evidence that training STEM graduate students with the same professional development programs as faculty makes them more likely to use effective teaching practices in their own academic careers (Wyse *et al.*, 2014)). Another benefit to training GTAs is improving the quality of instruction received by undergraduates (UG). UG students who take classes where the instructor used elements of scientific teaching were more likely to stay in STEM fields (Freeman *et al.*, 2014). UG students who do not experience active learning are more likely to leave STEM fields (Dirks and Cunningham, 2006).

In the absence of pedagogical training, instructors tend to use familiar techniques that they were exposed to as UGs (Brownell and Tanner, 2012; Stains *et al.*, 2018). Most UGs admit that very few of their instructors used any form of active learning. In fact, most have been exposed only to straight lecture and this then becomes their fallback when it is their turn in a teaching role. Gormally *et al.* (2016) found that there was a gap between what GTAs do and what they know about using evidence-based teaching techniques. They found that GTAs were curious about ways to use inquiry methods while teaching, but were unsure how to recognize reformed teaching, when to use those methods or even if they were using the right methods (Gormally *et al.*, 2016). This issue is definitely not being addressed in the overall graduate education experience in STEM disciplines even though GTAs are widely utilized to teach laboratory sections and some lectures. (Wyse *et al.*, 2014).

Research and reports recommend that higher education institutions provide GTAs with training that will make them teaching-ready when they become faculty members (Schussler *et al.*, 2015). Some studies indicate that GTAs may prefer one teaching style over another which leads to different teaching experiences than their faculty counterparts (Kendall and Schussler, 2012; Reeves *et al.*, 2016) However, none address a way to use specific measures that allow the GTA to be treated in the same manner as a faculty member and in

this way know what to expect if they decide to become faculty themselves (Schussler *et al.*, 2015)

In 2008, a discipline-based education researcher (DBER) was hired in the Department of Biology at WVU. One of the duties for this position was to reform the teaching of the freshman year biology curriculum. This effort took four forms: 1) pedagogical training for local faculty using a format adapted from the National Academies Summer Institute (SI) held at University of Wisconsin-Madison; 2) curricular reform of the introductory lectures using backward design, active learning and evaluation using validated concept inventories; 3) curricular reform in the introductory teaching laboratories, using backward design, guided and open-inquiry experiences and evaluation using validated concept-inventories; and 4) pedagogical training for GTAs who taught the laboratories. This chapter examines the effectiveness of the GTA training program, which consisted of a combination of peer-evaluation, focus on student-centered practices, and a scientific teaching workshop.

METHODS

Study Participants

A two-course introductory biology series is the first installment in a set of four core courses required for undergraduate Biology majors at West Virginia University (WVU). Each of the two introductory courses is made up of a lecture course and a laboratory. The laboratories are taught by GTAs, primarily from the Department of Biology, but occasionally from other STEM departments at WVU. The participants in this study include GTAs who taught during fall semesters in the laboratory associated with BIOL 115, Principles of Biology, who were willing to allow their teaching videos (a part of program evaluation for the GTA training program) to be used as part of this research project.

Pedagogical Training

Prior to the start of the semester, GTAs attended two days of pre-semester orientation which included a 3-hour workshop on scientific teaching: backward design, active-learning and student-centered learning. These reformed teaching strategies were also modeled for the GTAs during the weekly preparatory sessions. During each two-hour session, GTAs were engaged with the course materials created using backward design (Wiggins and McTighe, 1998) and scientific teaching practices with two explicit goals: 1) to train the GTAs in the material and 2) to model how to implement the materials in an active, student-centered way. This included discussion and reflection on the rationale for the teaching approaches embedded in the content and instruction for the laboratory activity.

Evaluation

As part of program evaluation for the new GTA training program that supported the newly reformed teaching laboratories, GTAs were videotaped twice during the semester: within two weeks of the start of the semester (pre) and again within the last two weeks of the laboratory (post). Between the two video-recording sessions, GTAs took part in two peer-

review sessions inspired by the mentoring provided to post-doctoral fellows in the Faculty Institutes for Reformed Science Teaching IV (FIRST IV, Ebert-May *et al.*, 2015). The teaching feedback guide developed for that program by Dr. Jennifer Momsen (North Dakota State University) was used to provide feedback during the peer-review sessions. This rubric focuses specifically on student-centered practices. The first peer-review was performed by a scientific teaching veteran (graduate student or faculty member formally trained in scientific teaching at either the original or WVU regional National Academies Summer Institute (NASI)) and the second review was by another GTA in the course. Typically, the GTA and the reviewer would meet soon after the teaching session to discuss the feedback and provide the reviewee with the completed evaluation rubric. In an attempt to monitor the quality of the feedback given during these sessions, the reviewee filled out a questionnaire after the feedback session to describe the amount and quality of the feedback. For consistency, all of the pre-videos were completed in the same week, capturing the same teaching laboratory, as were the post-videos.

Teaching videos were recorded during Fall (BIOL 115) and Spring (BIOL 117) semesters. The videorecorder was placed in the back of the room in the most advantageous spot to capture the GTAs teaching and interacting with their students during laboratory sessions. The recording was transferred to a high capacity hard drive and secured prior to analysis. Prior to evaluation, videotapes were randomly assigned numbers as identifiers to allow for anonymity.

See Chapter 3 Methods for a description of the training and evaluation procedures for the GTA teaching videos. The only difference in the approach to evaluating teaching videos between the two projects is that the Classroom Observation Protocol for Undergraduate STEM (COPUS, Smith, 2014) was not used to evaluate GTA teaching videos. The COPUS developers expressly stated that COPUS was not appropriate for evaluation of teaching

laboratories as it was developed and validated for lecture classes.

As part of this research project, GTAs were asked to complete a survey consisting of researcher-designed questions to gather demographics and perceptions of the training program. Because the survey was for research purposes and not part of the program evaluation for the new GTA training program, participation in the survey was voluntary. This survey was administered using the web-based survey platform, Survey Monkey (SurveyMonkey.com).

Thematic Analysis

Direct observation of a GTA teaching a laboratory was completed using the Teaching Feedback Guide adapted from J. Momsen in FIRST IV. The form has three questions to use to provide feedback: 1) Is there something the instructor is doing that the students could be doing? 2) What is the level of questions the instructor asks the students? and 3) Does the instructor facilitate the student process of learning activities? The guide also asked the observer to rate the level of student-centeredness they observe. Responses to these open-ended prompts were subjected to thematic analysis using a grounded theory approach (Braun and Clarke, 2006; Regassa and Morrison-Shetlar, 2009). Raters reviewed the responses to develop response codes. Codes and emergent themes were then discussed and refined. Responses were then binned into similar themes and a frequency for each theme or sub-theme was calculated as a percentage of total codes in the bin.

Survey

A simple survey was designed by the researcher to gather self-report data from GTA participants. The survey collected demographic and perceptions of the training using active learning techniques. GTAs were asked to complete this survey at the end of the semester following the post videotaping. Twelve GTAs completed the survey. The survey was administered online using Survey Monkey (www.surveymonkey.com).

Data Analysis

Data from the survey included frequency and descriptive statistics. The self-report data were summarized either by averaging the numerical score for each question or by calculating the average percentage response rate for each answer. The Kruskal-Wallis test was used to compare independent populations. The Wilcoxon signed rank test was used to make pairwise comparisons between dependent samples, such as pre-/post-responses. Data were collected from GTA teaching videos using two instruments; RTOP and timing analysis, described above. For each measure, descriptive statistics were carried out to calculate the mean and standard error when appropriate. Given the bounded nature of RTOP (0-100 scale) and timing analysis (percentages) scores, non-parametric statistical tests were used on analysis of variance across multiple samples (Kruskal-Wallis H test) and correlations between measures (Spearman rank correlation). All data analyses were carried out using the Statistical Package for Social Sciences (SPSS, version 25).

The protocol for this study was approved by the WVU Institutional Review Board (WVU IRB# H-24369).

RESULTS

The GTA training Program

A training program was developed for graduate teaching assistants (GTAs) in the Introductory Biology program in the Department of Biology at West Virginia University. The program was intended to better prepare GTAs for the newly reformed teaching laboratories. As part of a comprehensive Introductory Biology reform effort, the teaching laboratories had been transformed from recipe laboratories to guided- and open-inquiry experiences. The GTA training program used elements from the NASI@WVU and FIRST IV programs. Specifically, a scientific teaching workshop from the NASI@WVU and a program of peer

evaluation from the FIRST IV program were combined in order to increase GTAs' use of and comfort with active, student-centered teaching. As part of the program, GTAs took part in: a) a 3-hour pre-semester workshop on Scientific Teaching (Handelsman *et al.*, 2007); b) weekly preparatory sessions where they were engaged with evidence-based teaching materials; and c) two rounds of peer-review using the teaching evaluation rubric developed for FIRST IV (Ebert-May, 2015). Teaching videos from a subset of the GTA cohort from the first three years of the program were included in this project. Fifty percent of the GTAs had paired video-recordings (Female = 46%, Male = 54%).

GTA attitudes and perceptions of teaching

Half of GTAs voluntarily completed a survey composed of researcher-generated questions. This cohort was split between females (46%) and males (54%). Of the GTAs that responded to the survey, 75% received their UG education at a university and 25% at a liberal arts college. GTAs reported a variety of ways that they had experienced pedagogical training (Figure 1). The most common way to learn about teaching was observing another person teaching. GTAs were able to choose all options of teaching they observed prior to their own teaching experience. When asked what percentage of their teaching time was interactive, GTAs reported an average split of 32.5% passive and 57.5% active (Pearson's chi-squared, $p = 0.007$, $p < 0.05$). When separated by gender, there was no significant difference in time spent in passive versus active modes of teaching (Kruskal Wallis $p > 0.05$). In addition, more GTAs reported that they "always"/"often" used group/cooperative learning, inquiry-based learning and open discussion when teaching in the introductory teaching laboratories (Figure 2).

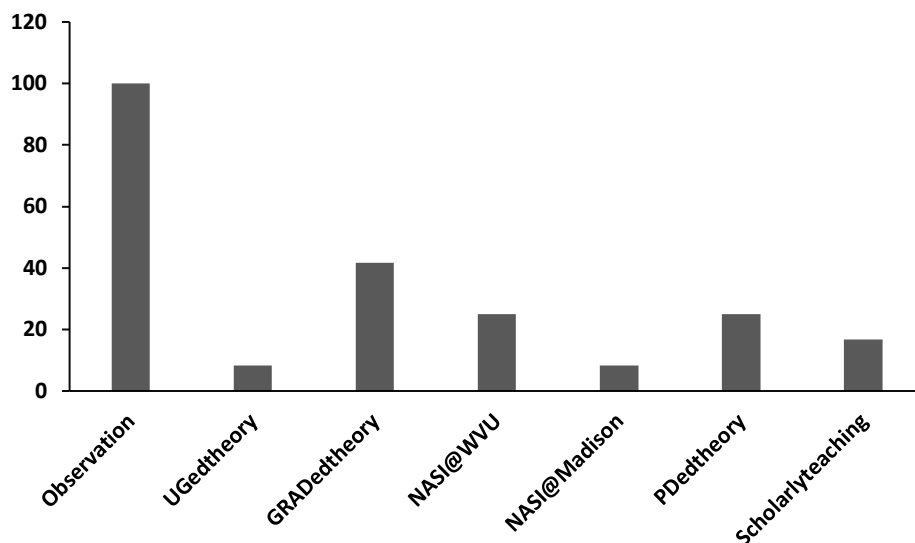


Figure 1. Types of training to which GTAs have been exposed . GTAs reported the type of teacher-education training that they have been exposed to. A majority of GTAs respond that the only training they have had is from observing (100%) instructors when they were UG. Respondents can choose more than one response type. Bars represent mean percentage of teaching type experienced. UGedtheory = Undergraduate education theory, GRADedtheory = Graduate education theory, PDedtheory = Professional development theory.

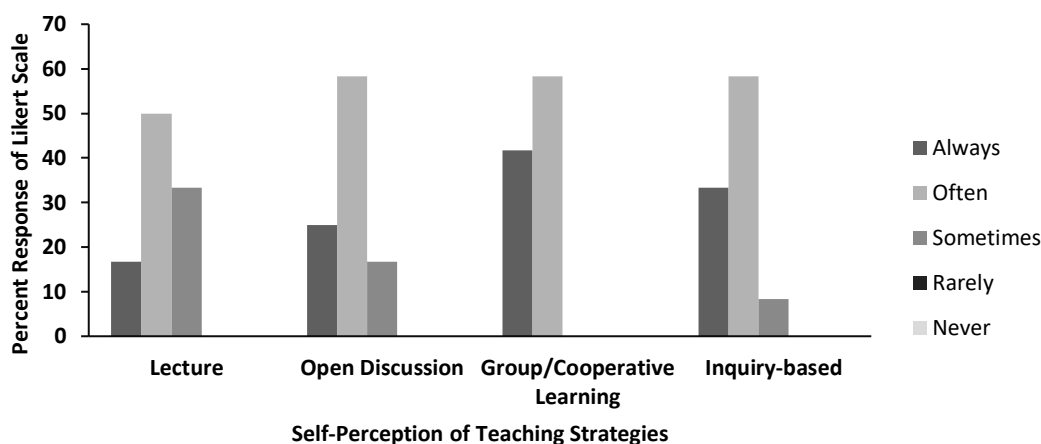


Figure 2. Perceived teaching strategies used during instruction. Response frequencies (Biology Education Pedagogy Survey) for the cohort year 2012 who described the various types of teaching strategies they perceived they use during a regular laboratory class instruction.

It is important that GTAs recognize student-centered techniques so they can employ them in their classes (Wyse *et al.*, 2014). GTAs identified the teaching strategies as student-centered with “students designing an experiment” identified by all GTAs, followed by “students guiding discourse”(83.3%), “completing an activity” (75%) and “clicker questions”

(58.3%) (Figure 3). GTAs responded to “how student-centered” they believe they were on a scale of 1- 10. Sixty-seven percent of GTAs believed they were 6 or higher on a scale of student-centeredness (mean = 6.2 +/- s.e. 0.5), while 58.3% of the respondents believed they were as student-centered as they would like to be.

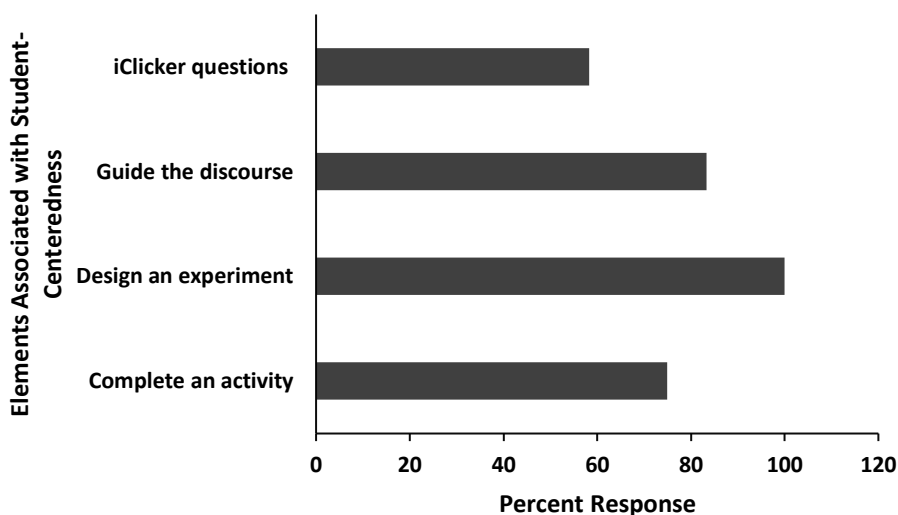


Figure 3. Elements associated with student-centeredness. All respondents perceive that Designing an experiment is an important element to have a student-centered classroom (100%). This was followed by students guiding the discourse in class (83.3%)

GTA Teaching Behaviors

GTAs teaching laboratories were video-recorded after having taken part in the pre-semester Scientific Teaching workshop but before taking part in two rounds of peer-review, first with a scientific teaching veteran (trained at the NASI at either University of Wisconsin or WVU) and then by another GTA. Analysis with RTOP (Sawada and Piburn, 2002) timing analysis metrics revealed the GTAs level of interactivity (percentage of time spent in an active mode calculated from the timing analysis rubric) and student-centeredness (RTOP score). There is a moderate positive relation between the percentage of time spent in an active mode and RTOP score before peer review (Figure 4, Spearman’s rho $r_s = 0.602$, $p < 0.05$) and a strong positive relationship between the two after peer-review (Figure 5, Spearman’s rho $r_s = 0.893$, $p < 0.000$). However, there was no significant difference between the amount of time

GTAs spent in an active mode before and after taking part in peer review and the same is true for their RTOP scores (Table 1. Wilcoxon signed rank test $p > 0.05$).

However, when compared to RTOP scores of faculty from other initiatives like the NASI@WVU or FIRST IV, the GTAs RTOP and Timing analysis scores were quite high both before and after peer-review. RTOP scores are separated into five categories (1 (15-30); 2 (31-45), 3 (46-60), 4 (61-75), 5 (>75)) that represent a transition from teacher-centered, categories 1 and 2, to increasing levels of student-centeredness, categories 3-5). The average RTOP scores for the GTAS put them in high Category 3 and low Category 4.

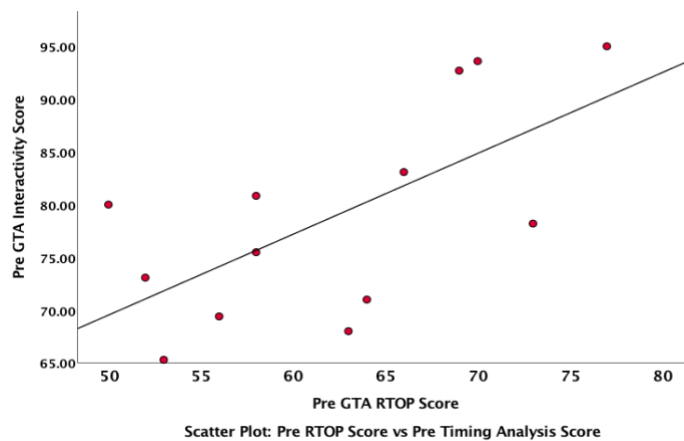


Figure 4. The relationship between RTOP and Interactivity Scores before peer review, where percentage of time in an active mode is calculated from timing analysis. Points are from all GTA teaching videos excluding those who did not have a paired vide, either pre or post. Spearman’s rho $r_s = 0.602$, $p < 0.05$.

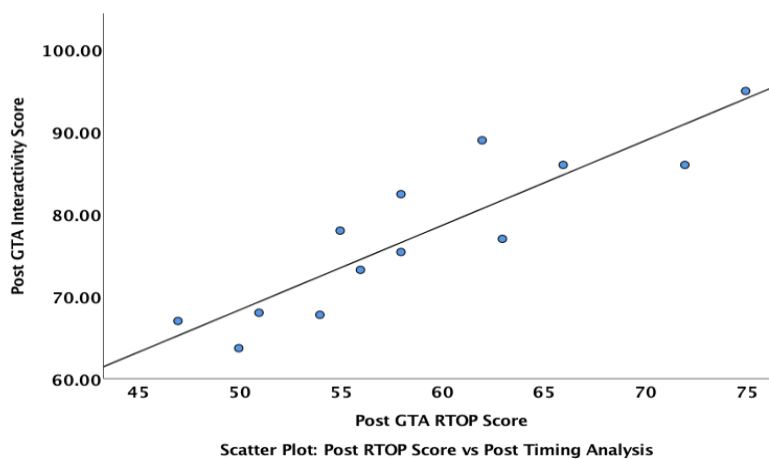


Figure 5. The relationship between RTOP and Interactivity Scores after peer review, where percentage of time in an active mode is calculated from timing analysis. Points are from all GTA teaching videos excluding those who did not have a paired vide, either pre or post. Spearman’s rho $r_s = 0.893$, $p < 0.000$.

Table 1. GTA RTOP scores before (pre) and after (Post) peer review.

Gender		Pre GTA RTOP Score	Post GTA RTOP Score	Pre GTA TA Score	Post GTA TA Score
Female	Mean	62.8	59.6	80.7	78.9
	N	6	6	6	6
	Std. Deviation	8.1	8.5	7.3	9.2
	Std. Error of Mean	3.3	3.4	2.9	3.7
Male	Mean	61.7	58.4	77.3	76.4
	N	7	7	7	7
	Std. Deviation	9.5	8.9	12.2	10.6
	Std. Error of Mean	3.6	3.4	4.6	4.0

Quality of feedback during peer review

At the end of each observation period, peer-observers prepared written comments that were provided to GTA instructors. The first round of observation (observer 1) was completed by a veteran observer. This person would have been trained at NASI@WVU or other comparable PD program. The second round of observation (observer 2) was completed by GTAs. The teacher feedback guide provides three questions to answer about general level of activity in the classroom that can be interpreted to determine student-centeredness. Figure 5 shows the bins that question 2 (What is the level of questions the instructor asks? Does it warrant class time?) was placed into to easily determine the level of questioning. Seventy-four percent of observer 1 thought the level of questions were appropriate or good and warranted the class time. At the second observation there was an increase in the appropriateness of questions asked (80%).

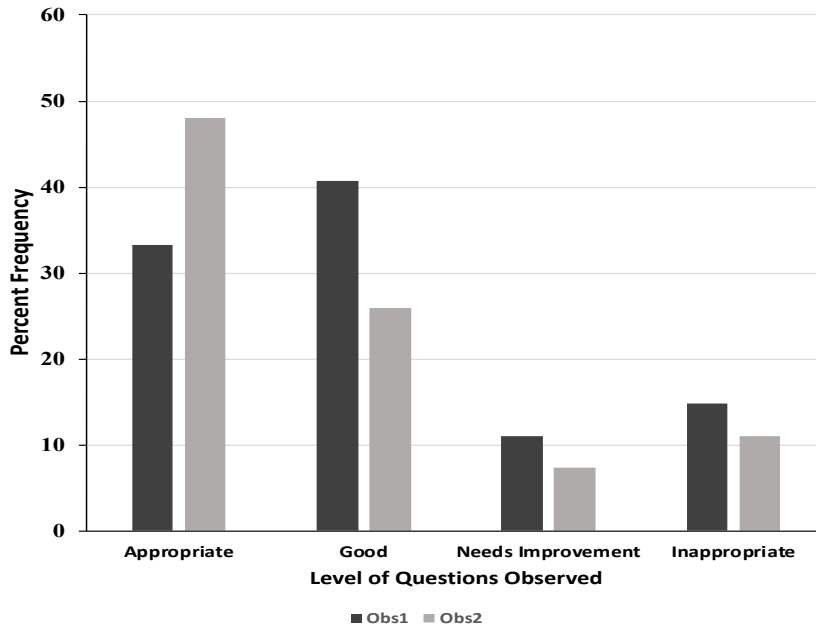


Figure 6. GTA responses to Question 2 from the feedback guide (“What is the level of questions the instructor asks? Does it warrant class time?”). Bars represent percentage average of frequency theme occurred across 27 peer reviews. Obs1 = First peer review by a Veteran observer with NASI@WVU training; Obs2 = Second peer review by a GTA or novice observer (has not attended NASI@WVU or similar program).

The trend was reversed for the third question (“Are GTAs meeting expectations to facilitate students during the learning process of active-learning lab activities?”) on the Teaching Feedback Guide (Figure 6). GTA were seen by the reviewers as facilitating student learning more often in the first peer review (81.5% of observations) than in the second (66.7%).

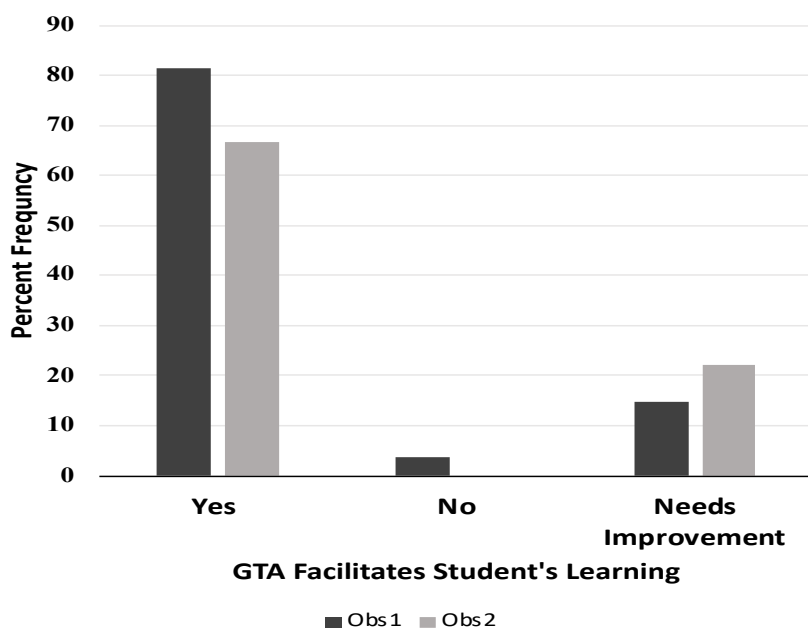


Figure 7. GTA responses to Question 3 from the feedback guide (“Are GTAs meeting expectations to facilitate students during the learning process of active-learning lab activities?”) Bars represent percentage average of frequency theme occurred across 27 peer reviews. Yes = GTAs are observed facilitating students during activities, No = GTAs are not acting as facilitators, Needs Improvement = there was some effort to let students answer questions on their own but GTAs would give more answers rather than let students find solution. Obs1 = Veteran observer with NASI@WVU training; Obs2 = GTA or novice observer (has not attended NASI@WVU or similar program)

DISCUSSION

The overall findings of the GTA assessment using scientific teaching elements in reformed laboratories indicate that GTAs are willing to make the effort to change towards more active teaching instruction. GTAs reported that some of their teaching techniques were based on directly observing instructors when they were UGs (Figure 1). What is encouraging is that 50% also reported having been taught graduate education theory techniques. This may be a result of attending courses offered through WVUs graduate education office for the Certificate in University Teaching. Thirty-three percent reported using professional development education theory and also 33% of the participants attended NASI@WVU.

On the research design survey, GTAs reported on average that on a scale of 1-10 (with ten being most student-centered) they averaged 6.2. One of the most important

components of this study was assessing GTA teaching in scientific teaching reformed laboratory sections on the RTOP metric similar to participants of NASI@WVU. The most concrete way to test the effectiveness of scientific teaching on an instructor is through assessment of their teaching. The premise is that GTAs will one day become faculty and using metrics like RTOP (Wyse *et al.*, 2014), will help them determine if they have made strides towards student-centeredness in a STEM classroom. Gormally *et al.* (2016) found a similar situation when they had TAs complete a pre/post survey. They found that TAs made some gains in conceptual understanding scientific teaching and inquiry (Gormally *et al.*, 2016). At the time of the pre-video recording, GTAs had only experienced the pre-semester scientific teaching workshop and a handful of preparatory classes. The RTOP scores and timing analysis had a moderate positive relationship based on the Spearman coefficient, showing that there is active-learning occurring during their teaching. Freeman *et al.* (2014), found that active-learning showed the most benefit in smaller class sizes, for which small laboratory sections would meet the criteria. In the post video recording the points are close together. In comparing the results from timing analysis and RTOP scores before (Figure 5), and after (Figure 6) peer review, the data from post-review videos congregate more closely and show a more distinct upward trend of higher RTOP scores and percent of activity during lab instruction. It should be noted that GTAs do not have control over the material that is used to teach lab sections. Consistency is maintained in the material since there are over 50 sections and all students need to gain the same skills. As such, GTAs have some built-in active aspects of the lab lesson plan; the non-lab portions are designed for them by a scientific teaching veteran to actively engage students. This may account for having rather high RTOP scores, category 4, distinguished by critiques and experimentation.

Concurrently, peer-review occurred before and after video recording of lab sessions. GTAs were observed by their fellow GTAs. From the theme analysis of observer notes, the

questions GTAs posed to their students became more appropriate from the first to second observations. However, there was a decrease in helping students meet the learning goals of making critical decisions for themselves. This may be the point in the semester where GTAs start to become overwhelmed with grading, their own course work and trying to meet their obligations. As such even if they are aware that they should be allowing students to formulate their own answers, hypotheses and designs, they may be giving answers to just get through a busy schedule. It may be that there was a positive effect of discussing ways to improve from a veteran GTA but not so much a novice GTA. Peer feedback is nonetheless important in building the confidence of GTA to use scientific teaching tools (Wood, 2009).

What the future holds for graduate student professional training programs

In their study Wyse *et al.* (2014) noted that the present level of training received by GTAs is insufficient. However, Wyse *et al.* (2014) did not have GTAs undergo peer-support and peer evaluations. They found that because GTAs had limited PDP, the RTOP scores from videotapes scored lower on the RTOP categories signifying less effective conceptual understanding of the content by students even when GTAs were experts in the content.

In our model, we proposed that the missing element for GTAs to move towards more student-centered instruction is peer evaluation. GTAs may become effective faculty members demonstrating best teaching practices in scientific teaching when peer evaluation is included. Ultimately, peer evaluation provided information that observation tools like RTOP did not fully describe. However, stress of the GTAs own coursework, needing to meet grading deadlines lessens the impact of peer evaluation and worrying about undergraduate perception of GTA teaching may (Reeves *et al.*, 2018) cause some GTA to not fully benefit from peer evaluation (Figure 7) Owens *et al.* (2018) determined that peer evaluation is a key factor for faculty members. Even faculty members who have had years to develop their teaching

portfolio feel it is one of the most important elements to be successful in active learning and scientific teaching (Owens *et al.* 2018).

Limitations

One issue, in hindsight, is that GTA participants should have taken the survey at the start of the semester to determine if there was actual movement in understanding that STEM courses should be taught with authentic science processes in mind. The answers given by GTA after the semester indicate that they are aware of these elements but the researcher is not sure if this is a cause of the scientific teaching workshop or prior exposure to other professional development.

References

American Association for the Advancement of Science. (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action, Final Report*. Washington, D.C.

Blouin, D. D., & Moss, A. M. (2015). Graduate Student Teacher Training: Still Relevant (and Missing?) 20 Years Later. *Teaching Sociology*, 43(2), 126-136.

Brownell, S. E., & Tanner, K. D. (2012). Barriers to Faculty Pedagogical Change: Lack of Training, Time, Incentives, and...Tensions with Professional Identity? *CBE - Life Sciences Education*, 11(4), 339–346.

Chen, X., Kimball, K., Frederick, J., & Graham, M. J. (2013). Training TAs in scientific teaching for the human physiology and anatomy laboratory. *Advances in Physiology Education*, 37(4): 436-439.

D'Avanzo, C. (2013). Post–Vision and Change: Do We Know How to Change? *CBE - Life Sciences Education*, 12, 373-382.

Dirks, C., & Cunningham, M. (2006). Enhancing Diversity in Science: Is Teaching Science Process Skills the Answer? *CBE - Life Sciences Education*, 5, 218-226.

Fagen, A. P., Crouch, C., & Mazur, E. (2002). Peer Instruction: Results from a Range of Classrooms. *The Physics Teacher*, 40(4): 206-209.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS*, 11(23): 8410-8415.

Gormally, C., Sullivan, C. S., & Szeinbaum, N. (2016). Uncovering Barriers to Teaching Assistants [TAs] Implementing Inquiry Teaching: Inconsistent Facilitation Techniques, Student Resistance, and Reluctance to Share Control over Learning with Students. *Journal of Microbiology & Biological Education*, 215-224.

Kendall, K., & Schussler, E. E. (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. *CBE - Life Sciences Education*, 11(2):187-99.

Ledbetter, M. L. (2012). Vision and Change in Undergraduate Education: A Call to Action Presentation to Faculty for Undergraduate Neuroscience. *Journal of Undergraduate Neuroscience Education*, 11(1), A22-A26.

Owens, M. T., Trujillo, G., Seidel, S. B., Harrison, C. D., Farrar, K. M., Benton, H. P., ... Tanner, K. D. (2018). Collectively Improving Our Teaching: Attempting Biology Department-wide Professional Development in Scientific Teaching. *CBE - Life Sciences Education*, 17 (1), ar2, 1-17.

Reeves, T. D., Hake, L. E., Chen, X., Frederick, J., Rudenga, K., Ludlow, L. H., & O'Connor, C. M. (2018). Does Context Matter? Convergent and Divergent Findings in the Cross-Institutional Evaluation of Graduate Teaching Assistant Professional Development Programs. *CBE - Life Sciences Education*, 17 (1), ar8, 1-13.

Stains, M., Harshman, J., Barker, M., Chasteen, S., Cole, R., DeChenne-Peters, S., ... Young, A. (2018) Anatomy of STEM teaching in North American universities. *Science*, 359: 1468

Wyse, S. A., Long, T. M., & Ebert-May, D. (2014). Teaching Assistant Professional Development in Biology Designed for and Driven by Multidimensional Data. *CBE Life Sciences Education*, 13(2), 212-223.

Chapter 5: General Discussion

History of STEM Education Reform

The idea that active learning improves student learning is not new. Studies addressing the use of active learning in STEM classroom are plentiful and spread across time. John Dewey (1916) set the stage for modern constructivism with his view that students were not “empty vessels” for teachers to fill with knowledge. He questioned the view that learning was passive. Much of today’s research about how students learn has its foundations in Dewey’s work on thought, experience and education (Dewey, 1910). In *Democracy and Education* (1916), Dewey makes the argument that to teach science in isolation does not fully serve the education of the student. It essentially disregards the students’ natural experiences outside the classroom (Dewey, 1916). Students are making connections, processing and solving problems in their natural world. When we fail to recognize students’ natural motives and curiosity about things around them, we underestimate what they are capable of understanding based on their own personal experiences (Dewey, 1916).

David Ausubel used Dewey’s theory as a foundation to build modern constructivism. His work clarified that students need to be actively engaged in building their own knowledge; knowledge cannot just be given to them (Ausubel, 1963, 2000). Ausubel’s theories deal with how students learn, their selectivity of what they learn and things they forget. He proposed that new learning tasks can be deciphered by using what the learner already knows, referred to as the learners “cognitive structure” (Ausubel, 2000). Students approach understanding new concepts by integrating them into their prior structure of knowledge (Ausubel, 2000). This theory gets at the root of student misconceptions. When students are passively exposed to new information, they may commit it to short-term memory without being required to merge it with their prior knowledge, leading to material that is easily forgotten (Fitzgerald &

Ausubel, 1963). When we do not use students' prior knowledge to enhance their learning early in their education, we end up with students who find post-secondary STEM education too difficult and leave for other disciplines (Freeman *et al.*, 2014).

It was not until the 1980-90s that multiple papers and reports started focusing on and calling for changes in post-secondary STEM (Angelo, 1998; Cross, 1990; Cross and Angelo, 1993; Cross and Steadman, 1996; Glassick, 1997; Huber 1999; NRC, 1983, 1999a, 1999b; Marzano, 1998; Rutherford, 1990). The National Academies and the National Research Council put efforts to reform teaching and learning in STEM on the center stage with several important reports. *A Nation at Risk* (1983) highlighted three major themes: 1) the intellectual rigor students experience was insufficient; 2) new curricular standards that promoted active student learning were needed; and 3) teaching needed to be evaluated more effectively. The last theme was particularly important as it would stimulate collection of data to show that current methods of teaching were primarily passive and lead to studies that would offer solutions for moving away from traditional lecturing.

In 1999, the *How People Learn* reports (NRC, 1999a, 1999b) provided a summary of what was known about human learning and cognition, followed by recommendations for how to bridge the gap between what was known and what was being practiced in the classroom, particular in college STEM. Still, most post-secondary STEM educators were not aware of this literature. Most had been well-trained in research, but received little to no training in teaching. This issue – having solutions from education literature that are not being implemented by the majority of practitioners – has posed a problem for post-secondary STEM education for decades. Physics was one of the earlier STEM disciplines that seemed to catch on that more needed to be done to bridge the gap between education literature and implementation in the classroom (Beichner and Saul, 2003; Mazur, 1996). Over several

decades physics education research has helped to shape the framework for research , development and dissemination of ways to improve physics teaching at the college level (Dancy and Henderson, 2009).

In 2003, the National Research Council helped get the national ball rolling for post-secondary biology education with *BIO2010: Transforming Undergraduate Education for Future Research Biologists* (NRC, 2003). This helped to examine all that had been done or proposed to bring STEM classes forward. STEM classes need to be more inquiry-based and active rather than just bombarding students with mass amounts of information that do not make any real-world connections. This report not only highlighted the message of reform to biology educators who were already paying attention to the education literature, but it also called for the creation of an immersive professional development workshop to train faculty who may not have been as aware of changes in STEM teaching and learning. This call was realized a year later in the National Academies Summer Institutes (NASI, Handelsman 2004, Pfund 2009). Prior to the formation of NASI there was much resistance to methods for transforming teacher-centered biology classes into more student-centered ones (Handelsman, 2004). The major tenet of scientific teaching, the curricular foundation of NASI, is that science education should be faithful to the true nature of science by capturing the process of discovery in the classroom (Handelsman *et al.*, 2004).

NASI trained many biology educators in evidence-based practices, gave us the scientific teaching approach and provided a model of successful professional development (Pfund, 1990) that others built-on:

- The American Association of Universities (AAU) STEM Initiative: AAU facilitates and collaborates with its member universities to promote initiatives that engage STEM undergraduates with evidence-based practices (AAU, 2017).

- Center for the Integration of Research, Teaching and Learning (CIRTL): a training program to help graduate students and post-doctoral fellows enter their first faculty positions already using evidence-based practices in their classes (CIRTL, 2006).
- Faculty Institutes for Reformed Science Teaching (FIRST IV) focuses its training programs on Post-doctoral fellows. Active learning encourages them to build inquiry lessons using reformed teaching strategies in active student learning (Gregg *et al.*, 2015; Ebert-May, 2015).

Like these endeavors, the NASI@WVU owes its origins to the original NASI.

NASI@WVU was the first full-length regional offshoot of the original NASI. It expanded the audience to include any STEM faculty interested in reformed teaching pedagogies (Bailey, chapters 2, 3). The results from the evaluation of the impact of NASI@WVU will contribute to the field in two important ways. The findings will contribute to our understanding of how professional development contributes to implementation of evidence-based practices. They will also provide evidence that professional development does contribute positively to education reform.

The argument has been made that professional development is ineffective at causing change in teaching practices, and the system of post-secondary STEM education needs a bigger “jolt” for more effective change to be seen (Wieman, 2012; Wilson, 2011). Charles Henderson’s research program has focused on the impact of professional development in physics and engineering for nearly a decade. He found that several professional development endeavors caused changes in motivation of participants to use the practices taught by the workshop and in their knowledge and skills of the new strategies (Henderson *et al.*, 2011). However, he found that participants were not implementing to the level that they reported after returning to their home campuses. Diane Ebert-May’s evaluation of participants from the original NASI and from her early efforts with the Faculty Institutes for Reformed Science Teaching, FIRST I and FIRST II, produced similar findings, that participants were not implementing active, student-centered strategies to the level that they were reporting (Ebert-

May *et al.*, 2011). For some the interpretation of these studies was that professional development did not work. For others, it was that professional development solved part, but not all, of the problem of facilitating broad adoption of evidence-based practices (Borrego and Henderson, 2014).

The data in chapters 2 and 3 of this dissertation support the latter interpretation by showing that effective professional development has the power to help participants increase their adoption of evidence-based practices. Not all of the NASI@WVU participants made it to the student-centeredness by definition of RTOP categories (Bailey, Ch. 3). The ones that did still had room for improvement. To expect a week of professional development to overcome the lack of awareness, knowledge and skills to use evidence-based practices and the challenges and barriers to implement that faculty face on their home campuses seems unrealistic. The more reasonable approach seems be to continue to develop and evaluate effective professional development, but to combine it with interventions that address the environment in which educators are attempting to reform their teaching (Borrego and Henderson, 2014).

Findings reported in chapter 4 demonstrate successes and areas of need in adapting professional development strategies for training graduate students. These data support the need for continued research to understand how to use what we know about successful professional development to develop effective pedagogical training for future faculty. Training future faculty to implement evidence-based practices before they enter their first position will help disrupt the pattern of producing an academic workforce that is prepared for only half of the expected duties of the job.

REFERENCES

American Associate for the Advancement of Science (2011). *“Vision and Change in Undergraduate Biology Education: A Call to Action.”* Washington, DC.

Association of American Universities. (2017). *Progress Toward Achieving Systemic Change. A Five-Year Status Report on the AAU Undergraduate STEM Education Initiative.* Washington, DC.

Anderson, R. (2002). Reforming Science Teaching: What Research says about Inquiry. *Journal of Science Teacher Education*, 13(1): 1-12.

Angelo, T.(1998). *Classroom assessment and research : An update on uses, approaches and research findings.* San Fransciso: Jossey-Bass.

Ausubel, D. (1963). *The psychology of meaningful verbal learning.* New York: Grune & Stratton.

Ausubel, D. (2000). *The acquisition and retention of knowledge: A cognitive view.* Boston: Kluwer Academic Publishers.

Biechner, R., & Saul, J. (2003). *Introduction to SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) project.* Proceedings of the International School of Physics. Varenna, Italy.

Bonwell, C., & Eison, J. (1991). *Active Learning; Creating Excitement in the Classroom.* ASHE-ERIC Higher Education Report No. 1. Washington, D.C.: The George Washington University, School of Education and Human Development.

Burrowes, P. A. (2003). A Student-Centered Approach to Teaching General Biology that Really Works: Lord's Constructivist Model Put to a Test. *The American Biology Teacher*, 491-494 & 496-502.

Angelo, T., & Cross, K. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.

Chickering, A., & Gamson, Z. (1987). Seven Principles For Good Practice in Undergraduate Education. *Washington Center News*. 1-7.

Crews, T., Ducate, L., Rathel, J., Heid, K., & Bishoff, S. (2011). *Clickers in the Classroom Transforming Students in to Active Learners*. Boulder, CO.:EDUCAUSE Center for Analysis and Research.

Cross, K. (1990). Teaching to improve learning. *Journal on Excellence in College Teaching*, 1:9-22.

Cross, K., & Steadman, M. (1996). *Classroom research: Implementing the scholarship of teaching*. San Francisco: Jossey-Bass.

Dewey, J. (1910). *How We Think*. New York, NY: D.C. Heath & Co., Publishers.

Dewey, J. (1916). *Democracy and education*. New York: MacMillan, The Free Press.

Fitzgerald, D., & Ausubel, D. (1963). Cognitive Versus Affective Factors in the Learning and Retention of Controversial Material. *Journal of Educational Psychology*, 54, (2), 73-84.

Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS*, 111 (23):8410-8415)

Glassick, C., Humber, M., & Maeroff, G. (1997). *Scholarship assessed: Evaluation of the professoriate*. San Francisco: Jossey-Bass.

Hestenes, D., Wells, M., & Swachamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30:141-158

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... Wood, W. (2004). Scientific Teaching. *Science* 304: 521-522.

Huber, M. (1999). *Developing discourse communities around the scholarship of teaching*. Paper presented at the annual meeting of the American Association of Higher Education, Washington, DC.

Marzano, R. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, CO: Mid-continental Regional Educational Laboratory (McREL).

Mazur, E. (1996). *Peer interaction, a user's manual*. Upper Saddle River, NJ: Prentice Hall.

National Committee on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, D.C.: United States Department of Education.

National Research Council (1999a). *How People Learn: Brain, mind experience and school*. Commission on Behavior and Social Sciences and Education, Washington, DC: National Academies Press.

National Research Council (1999b). *How People Learn: Bridging research and practice*. Commission on Behavior and Social Sciences and Education, Washington, DC: National Academies Press.

National Research Council. (2003). *Bio2010: Transforming undergraduate education for future research biologists*. Washington, DC: Nation Academy press.

Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D., ...Handelsman, J. (2009). Summer institute to improve university science teaching. *Science*, 324: 470.

Rutherford, F. (1990). *Science for all Americans*. New York: Oxford University Press.

Schroeder, C., Scott, T., Tolson, H., Huan, T., & Lee, Y. (2007). A meta-analysis of national research: effects of teaching strategies on student achievements in science in the United States. *Journal of Research in Science Teaching*, 44:1436-60.

Timmerman, B., Strickland, D., & Carstensen, S. (2008). Curricular reform and inquiry in biology: where are our efforts most fruitfully invested? *Integrative and Comparative Biology*, 48 (2) 222-240.

Wenderoth, M. (2007). A manual for the Scientific (Teaching) Revolution. *CBE - Life Sciences Education*, 6 (4): 271-272.

Wieman, C. (2012) Applying New Research to Improve Science Education. *Issues in Science and Technology*, 29 (1).

Wilson, S. M. (2011) How can we improve teacher quality? *Phi Delta Kappan*, 93 (2), 64-67.