Creative Science Through Inquiry: Improving Teacher Self-efficacy and Outcome Expectancy Through Adaptable, Mystery-based Professional Development

Shannon R. Sharp[a], George L. Rutherford, II[a] and Katherine I. Echols[b]

Corresponding author: Shannon R. Sharp (srsharp1@olemiss.edu)
[a]Center for Research Evaluation, The University of Mississippi, University, MS 38677
[b]Office of Research and Economic Development, Mississippi State University, Mississippi State, MS 39762

Keywords: professional development, science teacher education, problem-based teacher education, middle grades teacher education, self-efficacy, outcome expectancy

Abstract

Teacher self-efficacy is an important contributor to student outcomes, school climate and teacher retention. Outcome expectancy, a construct studied more commonly in health- and behaviour-related fields, may also positively impact school-related outcomes. Research shows that professional development can increase teacher confidence, but few studies have considered this connection for science-focused professional development, specifically. Our study assesses the impact of a science-focused, mystery-based professional development workshop for upper-elementary to high-school teachers. The hands-on, collaborative nature of this workshop allowed for generalisability to classrooms of various ability levels. Using the Science Teaching Efficacy Belief Instrument-A (STEBI-A) as a measure of science-teaching self-efficacy and outcome expectancy, we found that participants' self-reported self-efficacy and outcome expectancy significantly increased ($p < .001$ for each) over the course of the two-week workshop. This outcome is especially relevant to schools and districts interested in improving student outcomes, school climate and teacher retention rates.

Background

The Creative Science through Inquiry (CSI) summer workshop served as an outreach component for two National Science Foundation (NSF) Established Program to Stimulate Competitive Research (EPSCoR) awards (RII Track 1: OIA 0903787 and RII Track 2: OIA-1539035) and one NSF Career Award (CBET 1752036). Its purpose was to provide teachers with a hands-on opportunity to learn a variety of unique and interesting laboratory-based activities that they could easily adapt and use with their students. In providing this unique professional development experience, we hoped to increase participants' science-teaching self-efficacy and outcome expectancy.

Self-efficacy and outcome expectancy

Our research assesses the extent to which participating teachers' confidence in their ability to teach science effectively (self-efficacy) and the likelihood that their teaching will produce positive student outcomes (outcome expectancy) increased as a result of the CSI workshop. Bandura (1977) coined the constructs of self-efficacy and outcome expectancy, which have been used in an abundance of behavioural research. While outcome expectancy research tends to focus on health-related behaviours such as physical activity (Klusmann, Musculus, Sproesser & Renner, 2016; Williams, Anderson & Winett, 2005) or smoking cessation

(Kaufmann, Malloy & Haaga, 2020; Nikcevic et al., 2017) rather than occupational or educational ones, self-efficacy research is well-established in the educational realm (among others), and is particularly relevant for the purposes of this study.

On a broad scale, research suggest that schools whose instructional staff exhibit high levels of self-efficacy have better professional culture and more educational effectiveness (Bray-Clark & Bates, 2003; Guskey & Passaro, 1994; Ross, Hogaboam-Gray & Hannay, 2001; Tschannen-Moran, Hoy & Hoy, 1998). Teachers with high self-efficacy experience renewed confidence, are motivated, and are more focused than teachers with low self-efficacy (Gibson & Dembo, 1984). They are also better at responding to stressful and challenging situations in the classroom (Bray-Clark & Bates, 2003). Not surprisingly, teachers with high self-efficacy also positively influence their students' knowledge acquisition and display increased resolve when helping those having difficulties, which improves student learning outcomes (Bray-Clark & Bates, 2003). In fact, Stohlmann, Moore and Roehrig (2012) Nadelson, Seifert, Moll and Coats (2012), Yoon, Duncan, Lee, Scarloss and Shapley (2007) noted that teacher self-efficacy and content knowledge are directly related to student performance. Teacher self-efficacy plays an important role in students' academic outcomes (Podell & Soodak, 1993; Darling-Hammond, 2000), as well as a teacher's willingness to implement new instructional ideas (Allinder, 1994; Mohamadi & Asadzadeh, 2012). Alternately, Woolfolk-Hoy, Hoy and Davis (2009) found, specifically for science teachers, that '*Teachers who lack confidence in their knowledge of science content and pedagogy tend to deemphasise or avoid science teaching*' (p. 632). In general, low self-efficacy leads to performance deterioration (Bandura, 1997).

Low self-efficacy among science teachers often results from personal lab experiences that were inadequate or ineffective in terms of preparing them for teaching such hands-on lessons in their own classrooms, especially to younger students (Schoon & Boone, 1998). Additionally, many describe their own science-class experiences in primarily negative terms, including, "scary," "meaningless," "unpleasant," "boring" and "waste of time" (Tosun, 2000), which further reduces their science teaching self-efficacy. This perpetuates the cycle, resulting in fact-based (rather than hands-on content-based) science teaching that continues to be inadequate and ineffective.

The negative impact of low self-efficacy on one's drive to teach is especially important considering current teacher shortages, as it could result in teachers leaving the profession. According to data from the Department of Education (Cross, 2017), Mississippi, Louisiana and Alabama, the three states targeted by our workshop, all have teacher shortages in STEM fields. This is especially problematic in Mississippi as the Mississippi Department of Education reports one out of every three school districts has a critical shortage (Wierman, 2019). A report from the Economic Policy Institute (Garcia & Weiss, 2019) indicated that teacher shortages have a negative effect on teachers, students, and the public education system as a whole. Further, they reported teachers in critical-shortage states do not receive the training necessary to be successful, especially early in their career, which has a negative effect on teachers' morale and self-efficacy. Another report by Shockley, Watlington, Guglielmino and Felsher (2006) found that St. Lucie County School District in Florida reported a teacher turnover rate of 16.4% each year. Shockley states:

> '*Due to enormous costs associated with teacher attrition, combined with massive projected teacher shortages it is imperative that school districts design and fund teacher induction and mentorship programs targeted to support and keep effective teachers in the classroom*' (p. 113).

Training has been shown to be important in increasing and sustaining self-efficacy among K-12 teachers (Farokhzadian, Sabzi & Shahrbabaki, 2018; Gardner, Glassmeyer & Worthy, 2019; Goodale, 2013; Ross & Bruce, 2007), thus resulting in higher school morale and better student outcomes.

Rationale

Considering the research on why science teachers often exhibit low self-efficacy (i.e., negative personal lab and science-class experiences that were inadequate or ineffective; Schoon & Boone, 1998; Tosun, 2000), it follows that science teacher self-efficacy could be improved by providing them with positive science class and lab experiences. However, there is limited literature about science-based professional development's impact on the self-efficacy and, to an even lesser degree outcome expectancy, specifically (Blonder, Benny & Jones, 2014). This study aims to fill that gap. In general, studies found that effective professional development should model effective practices and be sustained, collaborative, content-focused, and interactive (Darling-Hammond, Hyler, Gardner & Espinoza, 2017; Garet, Porter, Desimone, Birman & Yoon, 2001). Yoon et al. (2007) added that effective professional development must allow teachers multiple opportunities to practice what they have learned. Blank, De las Alas and Smith (2008) provided the following specific elements of successful professional development: (1) includes over 50 hours, (2) is continuous and (3) is aligned to the curriculum. With this and teacher need for positive science training experiences in mind, we designed and implemented the Creative Science through Inquiry (CSI) professional-development workshop for upper-elementary, middle-school, and high-school teachers, which included all of the important elements of successful professional development, while also offering a unique mystery-based focus (with hands-on labs that all involved some sort of 'mystery' or 'crime' to solve) for a fun, hands-on experience that reinforces inquiry-based learning.

The CSI workshop specifically focused on inquiry-based learning because, as indicated in the *National Science Standards* (NRC, 1996) and *A Framework for K-12 Science Education* (NRC, 2012), this type of instruction is so important within science curricula. Inquiry-based learning is purported to increase students' science knowledge through guided learning and practicing critical thinking skills, a claim that is backed up by a meta-analysis conducted by Firman, Ertikanto and Abdurrahman (2019) which found that inquiry-based learning practices across seven countries resulted in an average effect size in the medium range. Incidentally, the highest effect size (0.88) came from a study conducted in the United States. Consequently, the focus of our workshop was inquiry-based learning, with a twist of mystery imbedded in each lab, to increase interest and engagement while also modeling best-practices.

Summary

We developed and implemented the CSI workshop, offering positive science and lab experiences with adherence to successful professional development criteria and a focus on inquiry-based learning, in an effort to increase science teacher self-efficacy and outcome expectancy. It spanned 8-10 days, modeled effective practice, and was collaborative, interactive, and content-focused, with links to the local science education standards. The hands-on nature of the workshop gave participants an understanding of how to implement each lab with inquiry-based learning in mind, and feedback from other participants and leaders ensured that they could adapt them to each classroom's individual needs. Thus, our workshop is unique in that it incorporates all of the following components: (1) well over 50 hours of training, (2) collaborative, (3) interactive/hands-on, (4) inquiry-based, (5) science standards-based, and (6) adaptable by grade and ability level. While there are other professional development programs similar to ours in some ways (Lotter, Smiley, Thompson & Dickenson, 2016; Posnanski, 2002;

Sinclair, Naizer & Ledbetter, 2011; Knowles, 2017), we did not find any that contained all of these components and evaluated its ability to impact self-efficacy and outcome expectancy.

Methods

Sites

The CSI Workshop was held in the summer of each of the grants' programming years, from 2013 to 2019, with two workshops in the summers of 2015 and 2016. Workshop locations changed year to year to target a variety of geographical areas and connect teachers with local universities containing STEM resources. The project used laboratory space at the following universities: Jackson State University (2013); The University of Mississippi (2014); The University of Southern Mississippi, Gulf Coast (2015 & 2016); Mississippi State University (2016); Tulane University (2017); The University of Alabama (2018); and Delta State University (2015 & 2019).

Participants

Over the seven years, 126 upper-elementary (N = 10, 8%), middle-school (N = 77; 61%), and high-school (N = 39; 31%) teachers participated in the CSI Workshop. The number of participants per cohort ranged from 9 to 18, with a mean of 14 per cohort. Most participants were female (N = 109, 87%) and came from a variety of science-related teaching emphases, including general science, biology, chemistry, physical science, physics and anatomy, microbiology, environmental science, earth and space science, marine biology, zoology, forensics, STEM, physics. Some mathematics, social studies, Gifted Education and Special Education teachers also participated.

To recruit participants, we sent information about the workshop (i.e., flyers and emails) to all of the middle-school science and mathematics teachers and their principals from each of the school districts in the target counties. Because workshop locations covered various areas in the southeast, we were able to reach a wide range of teachers. Part of the incentive for participation was a $1,000 stipend, travel reimbursement if the teaching institution was outside a 50-mile radius from the workshop location, a meal allowance, Continuing Education Units (CEUs; for all locations except Louisiana due to logistical constraints), and supplies to help the teachers run the labs in their classrooms. To make it easier for teachers to attend, the CSI Workshop took place during the summer months when teachers typically did not have other school-related duties.

Though we prioritised middle-school science and mathematics teachers because the curriculum was primarily tied to the middle-school standards, we also accepted elementary- and high-school teachers, as well as middle-school teachers who did not teach science or mathematics. Because of the wide range of grade levels and subjects taught, we selected labs that were easily adaptable for younger or older students, as is detailed below.

Workshop components

The workshop consisted of ten full days (with the exception of the 2015 and 2016 Gulf Coast workshops, which were eight days each) over two weeks. In an effort to build teacher self-efficacy and outcome expectancy, the program provided hands-on training using a variety of mystery-based laboratory activities that build students' inquiry skills while strengthening their content knowledge. The mystery component, which involved some type of unknown that was made into a 'crime' or 'mystery' that needed solving, was intended to increase students' interest and inquiry skills. The Education and Outreach Coordinator for the EPSCoR awards coordinated the workshops while a middle-school science teacher facilitated the daily activities.

We arranged the teachers in pairs for the entire workshop using an activity that required the use of deductive and inductive reasoning skills. Each day the teachers worked with their partner through several labs (described in more detail below) from start to finish, so they could understand the purpose, procedures, safety requirements and materials necessary. The workshop leader introduced a scientific concept and lab activity to the group and then the teachers worked through the lab with their partners, in the same way they would expect of their students. While they worked to complete the labs, we asked the teachers to think about the experience from their students' perspective. The workshop leader moved between pairs while they worked to address questions and concerns about both the scientific content and any reservations about implementation. After the teachers completed each lab, the workshop facilitator led the teachers in a discussion where they could share feedback and brainstorm ways to adapt the labs to their classroom based on their grade level and resources (e.g., space, time, and materials) available. Teachers also shared ideas and strategies for classroom-management practices when conducting labs, managing-administrator expectations, assessments, lab extensions and ways to collaborate with teachers in other subject areas to deepen interest and understanding. These discussions were an essential part of the workshop design and challenged the teachers to consider how to best incorporate inquiry, not just as a concept but within their specific circumstances.

Each year's workshop included at least 24 labs, most of which the workshop leaders adapted from kits they purchased from Flinn Scientific, a company that provides a wide variety of science lab kits for all grade levels. For curricular topics not available from Flinn Scientific, they found alternatives by searching Google or asking for resources from others familiar with the workshop. The leaders almost always adapted the labs they found to accomplish their primary goal of aligning the labs with the Mississippi Science Framework (Bounds et al., 2008), and then with the new Mississippi College- and Career-Readiness Standards (MSCCRS; Wright et al., 2016; see Table 1 for examples). This was an important aspect of the workshop because professional development is more successful in increasing self-efficacy when it aligns with the curriculum (Blank et al., 2008). We decided to use the Mississippi standards, specifically, because the majority of our participants were from Mississippi, and while each state has their own educational standards, this type of workshop can be replicated anywhere, keeping the local standards in mind.

Table 1. Example Lab Alignment with Science Standards

Lab Name	MSCCRS	Mississippi Science Framework
Build a Spectroscope/ Murder at Chem Fax Factory Lab	P.8.6, P.7.5A, P.7.5C, P.8.6	Competency 1-Inquiry, Competency 2-Physical Science
Dyeing for Forensics	L.6.1, L.8.2A, L.8.2B, L.8.2C, L.8.4B	Competency 1-Inquiry, Competency 2-Physical Science, Competency 3-Life Science

When selecting and adapting the labs to include in the workshop, workshop leaders considered the following components which they hoped would make it easier for teachers to use and students to learn from; thus, increasing the teachers' self-efficacy and outcome expectancy:

- **Lab procedures made sense and were easy to follow.** It was important that teachers felt comfortable using the labs we provided. If we found interesting labs that were not easy to implement or had difficult procedures, we often adapted them to make it easy for teachers to use, without compromising the lab's quality. We also edited the protocols year-to-year based on feedback from the teachers.
- **Lab materials were easy for teachers to obtain.** Because teachers often do not have a great deal of financial resources, we selected lab activities that would be inexpensive and used materials that could easily be obtained or were reusable for other labs. If the lab was not likely to be incorporated into the classroom due to a lack of resources, we did not use it.
- **Lab focus involved inquiry learning.** Some of the labs originally met the above criteria but failed to contain a robust inquiry element. When this was the case, we modified the lab to increase the focus on inquiry without compromising the content or quality of the lab. Ultimately, every lab contained some type of unknown that was made into a 'crime' or 'mystery' that needed solving.
- **Lab procedures could be easily adjusted to accommodate a variety of learners.** We strived to select labs that could be easily modified for different grade and ability levels. We believed that being able to use the labs, even if their students represented a different age/ability level than the lab was originally intended for, would positively influence the teachers' self-efficacy and outcome expectancy. Table 2 provides example descriptions of some labs used in the CSI Workshop, with possible adaptations.

Table 2. Example Lab Descriptions with Possible Adaptations

Lab	Build a Spectroscope/Murder at Chem Fax Factory Lab	Dying for Forensics
Description	Students build a working spectroscope using cardboard tubes and holographic diffraction grating. The spectroscopes are used to view and compare various light sources and collect data on the spectrums observed. They then solve a fictitious murder using the spectroscopes to differentiate line emission spectra observed during flame tests on various metals collected from suspects and compared to the evidence found on the victim.	Students use DNA fingerprinting to solve a mystery to identify the proper owner of a keepsake box that contains money and a lock of hair. Using simulated DNA samples from the hair and five individuals with potential claims to ownership, students will compare the samples using gel electrophoresis.
Potential Elementary Science Standards Adaptations	**4th Grade (standard P.4.6):** Have all light sources being observed be white; have light strike objects to observe behaviour (reflection, refraction, absorption); have light strike transparent, translucent, and opaque materials to observe behaviour. Compare the visibility of the metals, the colours emitted by the flame test, and the line spectrum visible through the spectroscope to explain how the visibility of an object is related to light.	**3rd Grade (standard L.3.2):** While 3rd grade students might not be able to load the gels, the samples being tested could be modified (along with the story) and the students just 'read' the results. For example, dye colours could be used to represent parents and the hair being tested would show a combination of the colours that match the correct parent. This would be a visual way to demonstrate heredity and introduce students to biotechnology.
Potential High-School Science Standards Adaptations	**Chemistry (standard CHE.7):** Have light pass through various gases to observe differences; adjust pressure, volume, temperature of a gas and use the light to make predictions about the behaviour of the gases under those circumstances. **Physics (standard PHY.4):** Have students calculate wavelengths or angles of the spectrograph. Have students use the absorption spectra to identify a mixture of substances used for the unknown sample.	**Physical Science (PHS.1):** Have the students calculate the molecular weight of the samples

Six of the nine workshops also incorporated tours of faculty labs at the university at which the workshop was held in an effort to expand the teachers' awareness of STEM opportunities available to students and encourage networking and resource-sharing between the university and workshop participants.

Additionally, upon completion of the workshop, we gave all participants a manual with details about all the labs they completed during the workshop, as well as a class set of materials needed to implement the some of the labs in their classrooms.

Data collection

The workshop facilitator administered paper forms of the Science Teaching Efficacy Belief Instrument-A (STEBI-A) to all participants on the first (pre-test) and final (post-test) days of each year's workshop. The STEBI-A is a 25-item assessment developed by Riggs and Knochs (1990) that measures two distinct factors associated with in-service teachers' science teaching beliefs: (1) *self-efficacy* (i.e., 'their own ability to perform the behaviours,' p. 626) measured via the Personal Science Teaching Efficacy Belief (PSTE) subscale and (2) *outcome expectancy* (i.e., 'expect[ing] certain behaviours to produce desirable outcomes,' p. 626) measured via the Science Teaching Outcome Expectancy (STOE) subscale. Initial (Riggs & Knochs, 1990) and updated (Bleicher, 2004) factor analyses indicate that the PSTE and STOE subscales represent constructs that are related but independent.

Constructs related to expected or confirmed relationships to teaching self-efficacy beliefs were all significantly and positively correlated to one or both STEBI scales, suggesting good construct validity. Reliability for the STEBI is also strong, with Cronbach alpha of .92 for the PSTE subscale and .77 for the STOE subscale. Further supporting its use for this study, the STEBI has been cited in multiple studies (Deehan, 2017), and though originally developed for elementary teachers, Moslemi and Mousavi (2019) determined its reliability and validity for secondary-school teachers as well (Cronbach's alphas of .89 and .72. for the self-efficacy and outcome expectancy scales, respectively).

We scored the STEBI according to the protocol set by its authors, assigning the following points for each item: 5 = Strongly Agree, 4 = Agree, 3 = Uncertain, 2 = Disagree and 1 = Strongly Disagree. We then followed the recoding rules, reversing scores such that 5 = Strongly Disagree and 1 = Strongly Agree for items that participants see in reverse or negative form. Based on this process, we generated sub-scores in self-efficacy and outcome expectancy beliefs for each form. The highest possible sub-score for self-efficacy is 65 (13 items) and for outcome expectancy is 60 (12 items).

In addition to the STEBI, participants completed an anonymous paper survey on the final day of the workshop. This survey was developed by workshop leaders and included an open-ended response item, "A thing of value…" to help us gauge which workshop components were considered most valuable to participating teachers, and thus most likely to result in feelings of teaching self-efficacy.

All testing and survey procedures were reviewed by Mississippi State University's Institutional Review Board (IRB).

Data analysis and findings

We used *IBM SPSS Statistics for Windows, Version 25* (IBM Corp., 2017) software to conduct paired-sample t-tests on the pre- and post-test STEBI scores for CSI Workshop participants across the seven programming years (N = 115; we did not include participants that did not complete the full pre- and/or post-test) to determine whether they increased their science teaching self-efficacy and outcome expectancy beliefs over the two-week workshop period. The results indicated that participants significantly increased their self-efficacy (t(114) = -4.910, p < .001) and outcome expectancy (t(114) = -10.986, p < .001; see Table 3). Differences in self-efficacy and outcome expectancy from pre-to-post represent medium effect sizes (*d* = 0.40 and *d*=0.67, respectively) providing evidence that participants felt the workshop had good and easily perceptible effects on (1) their capacity to teach science well and (2) positively impact their students' achievement through their teaching.

Table 3. STEBI Assessment Results (N = 115)

Sub-Score Area	Pre-Workshop Mean (SD)	Post-Workshop Mean (SD)	Mean Increase (SD)	t-test Result	p Value
Self-Efficacy	53.4 (5.6)	56.9 (4.8)	3.5 (3.4)	11.0	< .001*
Outcome Expectancy	43.3 (5.2)	45.5 (5.8)	2.2 (4.8)	4.9	< .001*

*Significant: p<.001

We also conducted non-parametric tests as a sensitivity check on these findings. To this end, we ran two Wilcoxon signed-rank tests comparing pre- and post-scores for Self-Efficacy and Outcome Expectancy. These non-parametric tests indicated similar findings – significant changes in self-reported Self-Efficacy (Z = 7.97, p < .001) and Outcome Expectancy (Z = 4.51, p < .001).

To analyze qualitative data collected from the survey item, "A thing of value…," we utilized descriptive coding (Miles, Huberman, & Saldana, 2013) to sort the responses into codes based on workshop components and common respondent themes. We then calculated the percentage of respondents who endorsed each theme by dividing the number of participants that endorsed the theme by the total number of meaningful item responses. Table 4 summarises the most-endorsed (i.e., endorsed by at least 10% of respondents) workshop components of value, which likely contribute to their increased self-efficacy.

Table 4. Most Endorsed Components of Value (N = 104)

Component	Endorsement level % (n)	Sample quote
Resources/materials provided	31% (32)	"New resources available for me and my students to use."
Relevant, adaptable activities/labs	22% (23)	"The use of practical examples that can be modified and used in any classroom."
Collaboration and networking for continued support	17% (18)	"Small groups allowed work to be done more efficiently and then we were able to collaborate as a whole groups [sic] at the end of the activity."
Hands-on practice	10 (10%)	"I really enjoyed being able to do a hands-on approach to learning versus sit and listen."
Ways to better-understand/help students	10 (10%)	"The experience of being a student again and being aware of the comfort level of my students."

Discussion

There is limited research on science-based professional development and its impact on self-efficacy and outcome expectancy, especially for programs where the content is adaptable to various grade and ability levels. To fill this gap, we studied the effects of a two-week science-based professional-development workshop on self-efficacy and outcome expectancy among upper-elementary, middle-school, and high-school teachers. Our results indicated that participating teachers' self-efficacy and outcome expectancy improved significantly from pre- to post-test ($p < .001$).

We propose that the professional-development workshop was successful because it contained components of successful professional-development models addressed in the literature. First, our workshop was conducted over a two-week period that exceeded the 50-hour threshold that literature suggests is needed for an impactful professional-development experience. It was also aligned to the science curriculum standards and encouraged collaboration and feedback from peers and leaders to devise suitable adaptations for various age and ability groups. Further, the workshop incorporated hands-on labs, allowing participants the opportunity to practice what they learned and experience the labs as their students would. Lastly, a particularly unique feature of this workshop was the mystery-based focus that encouraged learning through inquiry in a fun and engaging way for both the teachers and their students. Participants further indicate that some of these components, including adaptability of labs, collaboration and hands-on practice are among the most valuable workshop components. In addition to these, they also found the resources provided to be extremely valuable. Furthermore, we believe that because participation was voluntary and offered in a small-group format, participants were more engaged and benefited from more collaboration and individualised attention.

Our research is particularly important considering the STEM teacher shortages in our target states of Mississippi, Louisiana and Alabama, as the literature shows that teachers with low self-efficacy lack confidence in teaching science and avoid teaching it, which could worsen these shortages. This is important, as Garcia and Weiss (2019) note, because teacher shortages have a negative effect on teachers, students, and school systems as a whole. Self-efficacy is an important element when considering teacher retention because teachers with high self-efficacy have more confidence, motivation, and greater ability to respond to stressful situations in the classroom, and thus are more likely to remain in the field.

Limitations and future research

In an effort to provide lab experiences that could easily be adapted for various grade and ability levels, we sometimes changed which labs were offered from cohort to cohort. We considered feedback from previous participants regarding which labs were most easily generalised to their classrooms, which sometimes led us to remove previously-included labs and add more appropriate ones for future years. Additionally, as previously noted, we often changed the lab procedures from what was included in a kit or found online in an effort to include the mystery component, use less teacher resources, and/or be more appropriate for other grade or ability levels. Thus, because we did not have one consistent set of labs and procedures throughout the workshop years, replicability of the exact study design may be more difficult for those attempting to offer a similar professional development experience. For those wishing to do so, we recommend following the lab selection and adaptation criteria we explained above when considering which labs to include.

Despite our focus on specific criteria when selecting the labs (i.e., easy to follow lab procedures, easy to obtain materials, inquiry-based focus, and lab procedures that accommodate a variety of learners), not all participants could use all of the labs presented at each workshop once they returned to their classroom. This was sometimes the case for labs that required more resources than their school could provide or was provided by workshop leaders, or labs that were not as easy to adapt as some of the others. As noted below, changes in science standards also made it difficult for some teachers to fit the labs into the new curriculum. The participants were, however, encouraged to keep in touch with others in their cohort to continue to discuss ways to make each lab more useful for all participants. Many stated, through survey responses, that they intended to do this.

Another challenge we faced in the later years of the workshop involved changes made to the Mississippi science standards (as previously mentioned in the Methods section). 'Inquiry' was a separate standard in Mississippi when the workshop first began, and at that time the life, physical, and earth science standards were similar across grades, varying mainly in regards to depth of the information. The Mississippi standards have since changed and the inquiry component was dissolved into the other content standards which were restructured to focus on specific concepts for specific grades. This created a situation where some of the labs we used were no longer practical for all of the teachers in the workshop. More importantly, however, they learned how to create inquiry opportunities and adapt labs that were a more natural fit in their curriculum, despite the changes in standards.

We chose to include data from all participants in our analyses in an effort to maximise the sample size, and thus, the test's likelihood of detecting meaningful differences. Aggregate analyses such as these do come with limitations, however. For example, as we have noted previously, the number of hours and labs, as well as lab content, varied slightly from year to year, (though the minimum of 50 hours, 24 labs and standards-based mystery-involved content remained constant). Due to these variations, as well as a range of grade levels taught by participants, more-specific analyses (i.e., by cohort or by grade level taught) would likely yield more-specific, but potentially less- robust, results than those obtainable with a larger, aggregate sample.

In order to offer the best opportunity for positive change, we designed our workshop to incorporate research-based best practices with a positive and fun science-training experience. It is possible, however, that similar results could be achieved with a more-targeted mix of fewer components. Later research could explore this further, potentially focusing on the components identified as most valuable to the workshop participants, or a dose-response approach comparing various component combinations.

Acknowledgements

Research reported in this publication was supported by the National Science Foundation (NSF) under award numbers [OIA 0903787, OIA 1539035 and CBET 1752036].

References

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17,* 86-95.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1997). *Self-efficacy the exercise of control*. Englewood Cliffs, NJ: Prentice-Hall.

Blank, R. K., De las Alas, N., & Smith, C. (2008). *Does teacher professional development have effects on teaching and learning?: Analysis of evaluation findings from programs for mathematics and science teachers in 14 states.* Council of Chief States School Officers.

Bleicher, R.E. (2004). Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers. *School Science and Mathematics, 104*(8), 383-391.

Blonder, R., Benny, N., & Jones, M. (2014). Teaching self-efficacy of science teachers. In Evans, J., Luft, J., & Czerniak, C.M., & Pea, C. (Ed), *The role of science teachers' beliefs in international classrooms,* 3-16.

Bounds, H. M., Sewell, B., Kaase, K., Green, T., Chapman, C., & Wroten, M. (2008). *2010 Mississippi Science Framework*. Unpublished document.

Bray-Clark, N., & Bates, R. (2003). Self-efficacy beliefs and teacher effectiveness: Implications for professional development. *The Professional Educator, 26*(1), 13-22.

Cross, F. (2017). *Teacher shortage areas 1990-1991 through 2017-2018*. United States Department of Education Office of Postsecondary Education. https://www2.ed.gov/about/offices/list/ope/pol/ateachershortageareasreport2017-18.pdf

Darling-Hammond, L., Hyler, M., Gardner, M., & Espinoza, D. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/Effective_Teacher_Professional_Development_REPORT.pdf

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8,* 1-46.

Deehan, J. (2017). *The science teaching efficacy belief instrument (STEBI A and B)*. Springer International Publishing. doi:10.1007/978-2-329-42465-1

Farokhzadian, J., Sabzi, A., & Shahrbabaki, P. (2018). Improving self-efficacy of teachers in schools: Results of health promotion program. *International Journal of Adolescent Medicine and Health, 1-9.*

Firman, M., Ertikanto, C., & Abdurrahman, A. (2019). Description of meta-analysis of inquiry-based learning of science in improving students' inquiry skill. *Journal of Physics Conference Series, 1151*(2). https://iopscience.iop.org/article/10.1088/1742-6596/1157/2/022018

Garcia, E., & Weiss, E. (2019, July). *The role of early career supports, continuous professional development, and learning communities in the teacher shortage: The fifth report in 'The perfect storm in the teacher labor market series*. Economic Policy Institute. https://www.epi.org/publication/teacher-shortage-professional-development-and-learning-communities/

Gardner, K., Glassmeyer, D., & Worthy, R. (2019). Impacts of STEM professional development on teachers' knowledge, self-efficacy, and practice. *Frontiers in Education, 4*(26), 1-10.

Garet, M. S., Porter, A.C., Desimone, L., Birman, B. F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569-582.

Goodale, T. (2013). The influence of a ROV themed engineering design workshop on in-service teacher self-efficacy. *Integrated STEM Education Conference,* 1-5.

Guskey, T. R., & Passaro, P.D. (1994). Teacher-Efficacy: A study of construct dimensions. *American Educational Research Journal, 31*(3), 627-643.

IBM Corp. (2017). IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp.

Kaufmann, A., Malloy, E. J., & Haaga, D. A. (2020). Examining outcome expectancies for smoking vs. abstinence among adult daily smokers. *Addictive Behaviors, 102*. 10610. Doi:10.1016/j.addbeh.2019.106140

Klusmann, V., Musculus, L., Sproesser, G., & Renner, B. (2016). Fulfilled emotional outcome expectancies enable successful adoption and maintenance of physical activity. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.01990

Knowles, J.G. (2017). *Impacts of professional development in integrated STEM education on teacher self-efficacy, outcome expectancy, and STEM awareness* (Publication No. 10270934). [Doctoral dissertation, Purdue University]. ProQuest Dissertations Publishing.

Lotter, C., Smiley, W., Thompson, S., & Dickenson, T. (2016). The impact of a professional development model on middle school science teachers' efficacy and implementation of inquiry. *International Journal of Science Education, 38*(18), 2712-2741.

Miles, M., Huberman, A., & Saldana, J. (2013). *Qualitative data analysis: A methods sourcebook*. Sage. Thousand Oaks, CA.

Mohamadi, F., & Asadzadeh, H. (2012). Teaching the mediating role of teachers' self-efficacy beliefs in the relationship between sources of efficacy information and student achievement. *Asia Pacific Education Review, 13,* 427-433.

Moslemi, N., & Mousavi, A. (2019). A psychometric re-examination of the Science Teaching Efficacy and Beliefs Instrument in Canadian context. *Education Sciences, 9*(1), 1-10.

Nadelson, L, Seifert, A. Moll, A., & Coats, B. (2012). I-Stem summer institute: An integrated approach to teacher professional development in STEM. *Journal of STEM Education, 13*(2), 69-83.
National Research Council. (1996). *National Science Education Standards.* National Academic Press.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* National Academic Press.
Nikčević, A. V., Alma, L., Marino, C., Kolubinski, D., Yilmaz-Samanci, A. E., Caselli, G., & Spada, M. M. (2017).
Modeling the contribution of negative affect, outcome expectancies and metacognitions to cigarette use and nicotine dependence. *Addictive Behaviors, 74*, 82-89.
Podell, D. M., & Soodak, L. C. (1993). Teacher efficacy and bias in special education referrals. *Journal of Educational Research, 86*(4), 247-253.
Posnanski, T. (2002). Professional development programs for elementary science teachers: An analysis of teacher self-efficacy beliefs and a professional development model. *Journal of Science Teacher Education, 13*(3), 189-220.
Riggs, I., & Knochs, L. (1990). Towards the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*, 625-637.
Ross, J., & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of randomized field trial. *The Journal of Educational Research, 101*(1), 50-60.
Ross, J., Hogaboam-Gray, A., & Hannay, L. (2001). Effects of teacher-efficacy on computer skills and computer cognitions of Canadian students in grades K-3. *The Elementary School Journal, 102*(2), 141-156.
Schoon, K. J., & Boone, W. J. (1998). Self-efficacy and alternative conceptions of science of preservice elementary teachers. *Science Education, 82*(5), 553–568**.**
Shockley, R., Watlington, E., Guglielmino, P., & Felsher, R. (2006). The high cost of leaving: An analysis of the cost of teacher turnover. http://www.pearsonschool.com/live/assets/200727/2006_11ShockleyGuglielminoWatlington_558_1.pdf.
Sinclair, B., Naizer, G., & Ledbetter, C. (2011). Observed implementation of a science professional development program for K-8 classrooms. *Journal of Science Teacher Education, 22*(7), 579-594.
Stohlmann, M., Moore, T.J., & Roehrig, G.H. (2012). Considerations for teaching integrated STEM education. *Journal of pre-college engineering educational research, 2*(1), 28-34.
Tosun, T. (2000). The beliefs of preservice elementary teachers toward science and science teaching. *School Science and Mathematics, 100*(7), 374–379.
Tschannen-Moran, M., Hoy, A., & Hoy, W. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.
Wierman, M. (2019, July 31). Addressing Teacher Shortage in Mississippi. *Edmentum*. https://blog.edmentum.com/addressing-teacher-shortage-mississippi
Williams, D., Anderson, E., & Winett, R. (2005). A review of the outcome expectancy construct in physical activity research. *Annals of behavioral medicine: A publication of the Society of Behavioral Medicine, 29*, 70-79.
Woolfolk-Hoy, A., Hoy, W., & Davis, H. (2009). Teachers' self-efficacy beliefs. In Wentzel, K.R. & Miele (Eds.), D.B. *Handbook of Motivation at School (Educational Psychology Handbook Series)* (pp. 627-653). Routledge/Taylor & Francis Group.
Wright, C. M., Benton, K. S., Massey, J., Oakley, N., Clemons, W., Smith, T., Davis, M., & Sampsell, J. (2016). *2018 Mississippi College- and Career-Readiness Standards for Science*. Unpublished document.
Yoon, K.S., Duncan, T., Lee, S., W. Y., Scarloss, B., & Shapley, K.L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement.* U.S. Department of Education, Institute of Education. Sciences. https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf