2024 LEAP Challenge

LEAP

A Plan to Trace Associations Between Practical Education Network and Student and Teacher Outcomes

Project Host:

Practical Education Network

Practical Education Network

Fellows:

Lindsay Kincaid, Team Lead, Social Entrepreneur Fellow

Oluwakemi Olurinola, Research Fellow

Animesh Priya, Social Entrepreneur Fellow

Teomara (Teya) Rutherford, Research Fellow

TABLE OF CONTENTS

Executive Summary

Introduction

In Ghana, students often describe their education as "Chew and pour, pass and forget," reflecting a system heavily reliant on rote memorization, particularly in STEM subjects. To address this, the Government of Ghana has set an ambitious goal of having 60% of students pursue STEM fields at the university level. However, this shift requires a transformation in how STEM education is delivered.

The Practical Education Network (PEN) is at the forefront of this change, empowering educators to create hands-on, inquiry-based learning environments using locally available, low-cost materials. PEN's innovative approach bridges the gap between theory and practice, enabling students to engage with STEM subjects in a tangible, interactive way. To date, PEN has trained around 150 trainers and 9,000 teachers, reaching 2 million learners who have benefited from meaningful, hands-on STEM education for the first time.

The LEAP project aimed to support PEN in three critical areas: designing a tracer study to evaluate long-term program impact, providing recommendations for organizational growth, and developing strategies to contextualize and communicate program effects effectively.

Organisation's Role & Strength

PEN's mission is to enable every African child to learn by doing. They are transforming STEM education in Ghana and beyond by equipping educators with the tools and training they need to implement practical, experiential learning. The organization's innovative teacher training program reduces barriers to hands-on learning in material resource-constrained settings, fostering critical thinking, creativity, and problem-solving skills among students.

PEN's impact extends beyond classrooms: its methods have been integrated into Ghana's national science curriculum and it has helped bring "hands-on" and "STEM" into everyday parlance amongst education stakeholders in Ghana. While the organization extends its reach geographically, it is also interested in understanding the depth of its impact on students. Specifically it seeks to uncover the extent to which students taught by teachers trained in hands-on approaches may be more likely to pursue STEM majors, which can directly inform policies to achieve the government's goals. PEN's emphasis on sustainability through the reliance on locally-available materials and on scalability through the training-of-trainers model positions its approach to be replicated across diverse educational contexts.

Need Summary

To achieve its mission, PEN seeks to deepen its impact through robust research, improved program fidelity, and effective communication of results. Key priorities include:

- **Evidence Generation**: Developing a tracer study to track the long-term impact of its teacher training programs on student outcomes and career choices.
- **Program Fidelity**: Strengthening implementation practices to ensure consistent quality across training sessions and classrooms.
- **Growth & Communication**: Expanding digital resource libraries, fostering community forums for teachers, and refining strategies to communicate program outcomes to stakeholders, including government officials and funders.

The LEAP project focused on addressing these priorities through tailored recommendations and actionable plans.

Solution Summary & Next Steps

The LEAP Fellows provided PEN with four key deliverables to support its goals:

1. **Tracer Study Plan**: Fellows designed a comprehensive tracer study to evaluate the long-term outcomes of PEN's programs on teachers and students. This included developing sampling procedures, measurement tools, and analysis strategies to assess teacher adoption of hands-on methods and student progression into STEM fields (Deliverable 1).
2. **Organizational Growth Recommendations**: Fellows proposed strategies to improve program fidelity, such as incorporating a conceptual framework for measuring implementation consistency. Additional recommendations included expanding PEN's digital resource library and fostering online community forums to facilitate peer learning among educators (Deliverable 2).
3. **Future Research Options**: Fellows outlined research opportunities to further PEN's understanding of the impact of their program. This included a detailed typology of studies to assess the impact of PEN's programs and strategies for aligning these studies with organizational priorities (Deliverable 3).
4. **Contextualizing & Communicating Effects**: Fellows provided a framework for effectively communicating PEN's impact to diverse audiences, including government stakeholders. This included recommendations on using standard deviations to present effect sizes and models for clear, impactful storytelling (Deliverable 4).

These deliverables collectively will equip PEN to enhance its program delivery, generate evidence of its impact, and scale its approach to transform STEM education across Ghana and beyond. The methodology developed through this project can also serve as a model for other organizations seeking to implement and evaluate hands-on STEM education programs in settings with minimal material resources.

Deliverable 1: Plan for Tracer Study

Justification of Need and Relevant Theory of Change

Based on the need to both better understand and promote its program, PEN had applied for small grants to support its "tracing" of students after they engaged with teachers trained through the PEN professional development (PD) program. One such small grant for $19K was awarded in October, 2024, and PEN had immediate need to develop a plan to carry out a *Tracer Study*. PEN presented this need to the LEAP fellows, who sought to understand what goals PEN felt would be accomplished from the study. The original plan presented to the LEAP fellows was to solicit names from a selection of teachers who had engaged with different PD models within PEN across years. One teacher from each year would participate and each would nominate students with the goal of PEN identifying five students for each teacher. Students would be invited to participate in a semi-structured interview to identify their STEM motivations, aspirations, intentions, and the factors from their earlier education and experiences to which they attributed these outcomes.

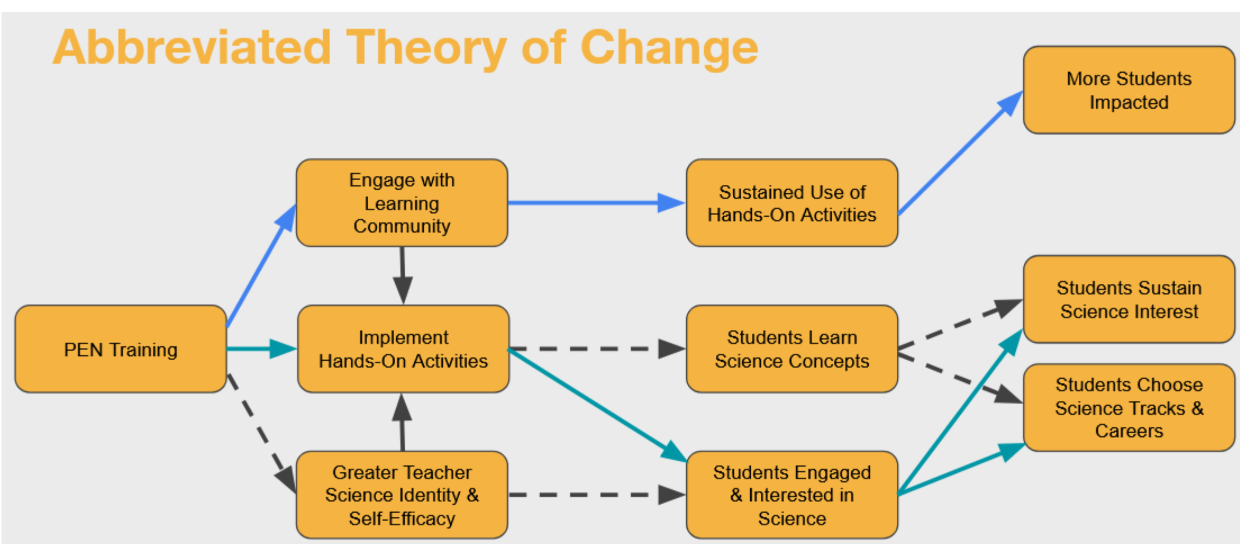
LEAP fellows reviewed the study plan and prior research PEN had conducted on its program outcomes. After meetings with PEN, LEAP fellows constructed an abbreviated theory of change (Figure 1) to represent PEN's goals and the hypothesized pathways through which teacher participation in PEN's program could lead to valued student outcomes. This theory of change is based both on PEN's expectations and on relevant scholarly research. In aligning the theory of change to the Tracer Study goals, the LEAP fellows focused on the factors that would support both teacher sustainability (tracing teachers) and student outcomes (tracing students).

The blue pathways within the theory of change focus on tracing teachers. PEN's training is expected to change teacher practice in meaningful ways. Sustainability of this change in practice is hypothesized to be supported by teacher engagement with a learning community. Research, such as Steyn (2017), collectively emphasizes the importance of collaborative learning communities and the integration of sustainability into professional development programs. The studies (Bendtsen et al., 2021; Steyn, 2017) highlight the need for grounding professional development in practical experiences. According to Bendtsen et al. (2021), "sustainable" in this context refers to the impact of professional learning beyond the duration of the course in question. Iversen and Dindler (2014) distinguished between four ideal typical forms of sustainability: maintaining, scaling, replicating, and evolving. These elements are crucial for fostering a sustainable community and ensuring the continued use of sustainable practices in education. These communities help sustain the use of new practices and skills gained during PD by providing ongoing support and encouragement (Meesuk, et al., 2021). Focusing on communities of peers ensures these community members can have a deep understanding of local issues and can offer valuable insight (Pandey & Kumar, 2019). PEN has available some mechanisms for supporting teacher community through a lightly-used WhatsApp group, follow up with teachers by PEN staff, and naturally-occuring teacher relationships among teachers. Whether and how often teachers engage with these communities and teachers' continued use of PEN's hands-on activities will be explored in the Tracer Study.

The teal colored pathways shown in Figure 1 are the motivational pathways through which participation in PEN hands-on science activities can influence students' subsequent motivation, engagement, and choice in science. Prior research conducted on an international sample as part of PISA has shown that frequency of exposure to hands-on science activities is associated with greater engagement and motivation toward science (Hampden-Thompson & Bennett, 2013). However, it is important that the activities be high-quality, as there is potential for some hands-on activities to have negative associations with interest (Holstermann et al., 2010). Hidi and Renninger's (2006) four-phase model of interest development describes how experiencing a situationally-interesting activity, such as the hands-on activities PEN teachers implement, can lead, over time, to a sustained personal interest in a subject or topic. These interests, and related motivational beliefs, such as stronger beliefs in their own science competence and greater value for science, such as finding science useful and important, should predict students' choice into more science activities, science tracks in school, and ultimately, science careers (Eccles & Wigfield, 2020). These pathways will be investigated in the Tracer Study by examining whether students recall engaging in hands-on activities, where they attended junior high school, and their current science motivation, intentions, and enrollment.

Figure 1

PEN Abbreviated Theory of Change



The dashed pathways in the theory of change involve constructs that will not be investigated in the Tracer study but that are articulated here for future investigation and to provide context for the Tracer Study investigations. The two lower pathways are from PEN training to teacher science self-efficacy. Self-efficacy is an individual's belief that they can engage in the behavior necessary to accomplish a specific task (Bandura, 1977). One important driver of self-efficacy is the individual's previously successful attempts at similar tasks (Bandura, 1977)—PEN's training includes sample activities and supports teachers in their execution of these activities. These experiences support positive self-efficacy beliefs. Teachers who feel more confident in their ability to execute tasks are more likely to implement them (Rutherford et al., 2018; Tschannen-Moran et al., 1998), thus positive teacher self-efficacy leads to greater implementation of hands-on activities, which in turn can support positive student motivation. There are likely other paths from stronger teacher self-efficacy in science to student motivation beliefs through additional pedagogical practices apart from hands-on activities. In addition to self-efficacy, PEN training, with its opportunities to practice hands-on science in a community, is likely to increase teachers' own science identity and their identity as inquiry-oriented teachers that value hands-on science activities (see Avraamidou, 2014; Luehmann, 2007). This identity is also likely to lead to implementation of more hands-on activities (e.g., Madden & Wiebe, 2013). Although many of these pathways are not part of the Tracer Study, the solid gray pathway from teacher science identity and self-efficacy to implementation of hands-on activities can be studied within the tracer study. Affirming this pathway will illuminate an important aspect of professional development to support hands-on activity implementation.

The dashed pathways in the middle of the model focus on student learning and knowledge. A deep literature supports links between active engagement and better student learning (Freeman et al., 2014). The development of knowledge of science concepts is important both for student ability to choose science tracks and careers and for further development of sustained interest (Hidi & Renninger, 2006).

Based on the Theory of Change articulated above and the goals of the Tracer Study as situated within the theory, the following research questions can be addressed in the Tracer Study:
1. Does student-reported hands-on learning in junior high school relate to STEM motivation and outcomes among senior high students? 2. Does enrollment in a classroom in which teachers had received PEN training relate to STEM motivation and outcomes among senior high students?

3. To what extent are teachers' use of hands-on activities maintained after the PEN training?
4. To what extent do teachers engage with PEN after their initial training?
5. To what extent does this engagement and other community support predict teachers' continued implementation and positive motivations?
6. To what extent does positive motivation predict continued implementation?

Sampling Plan and Procedures

Teacher Tracer

The Teacher Tracer Study is largely a descriptive study to answer research questions #s 3–6. All teachers who have participated in PEN training can be included in the Teacher Tracer Study. We suggest focusing on teachers who have been trained within the last three years so as to ensure that students who were trained by PEN teachers during JHS are still in SHS at the time of this study. This is approximately 500 teachers since 2021, distributed across multiple Regions of Ghana as well as Rwanda and Liberia. Teachers can be contacted by phone, by direct mailer, or by visit to their schools when possible/necessary. Teachers should be offered some small compensation for completing study activities.

All teachers included in the Teacher Tracer Study will complete a survey. A small sample of teachers will participate in a classroom observation (optional, by teacher volunteer). Those participating in the observation should be offered additional compensation.

In addition, the Teacher Tracer Study involves PEN's analysis of their own data over the approximately nine years of their training. PEN should prepare descriptive statistics on its contact with teachers: how often they engaged with the WhatsApp group, how often PEN reached out for follow up, and any other engagement. These numbers should also be matched with individual teachers when possible, to serve as explanatory variables for teacher outcomes (see measures and analysis sections, below).

PEN should aim for at least a 70% response rate in the survey sample and for 50% of that sample to be represented in the observation sample (funding permitting). This will permit a representative sample and avoid non-response bias.

Student Tracer

The Student Tracer Study is intended to illuminate pathways within the theory of change from engagement in hands-on activities in junior high school science to student motivation for science and their intentions and enrollment. These pathways align with research questions #1 and #2. Students should be recruited to complete the Tracer Study survey based on likely senior secondary pathways from junior secondary schools where PEN teachers have been trained. PEN should aim for a final sample size of at least 200 students who had previously attended at least 15 PEN schools and 15 non-PEN schools. These suggested numbers are just rules of thumb, as more formal power analyses would require typical Intraclass Correlations (ICCs) for the outcome variables among this population and average numbers of students per teacher or school.

We suggest that PEN offer some form of incentive to students for completing the survey (such as an entry into a drawing). Rewards should be resolved quickly given the propensity of students to change contact information frequently. If PEN wishes to maintain a database of survey respondents for future longitudinal or other follow-up surveys, we suggest obtaining multiple means of contact for each student respondent.

Measures

Teacher Measures

The teacher measures provided below are meant to encompass all the pathways described prior. PEN can select which of the constructs are most important for them to assess for the Tracer Study. As a rule of thumb, a 50-question Likert-type survey will take approximately 20 minutes to complete, depending on Internet speed. Before administering any surveys at scale, PEN should engage in *cognitive interviewing* for both the teacher and the student surveys (see Beatty & Willis, 2007; Rutherford et al., 2021). In cognitive interviews, individuals who are within the target group (e.g., teachers of certain grades in Ghana) are asked to go through the survey to note confusing wording or other issues they have completing it. Those administering the cognitive interview can either ask the respondent to think aloud while they complete the survey or can probe to ensure understanding of specific questions in line with construct definitions. Completing cognitive interviewing can help support construct validity, which is important in illustrating pathways within the theory of change.

Measures of teacher community support. Research on teacher implementation fidelity and intervention sustainability note the importance of community support to maintaining gains from professional development (Meesuk et al., 2021; Steyn, 2017). A series of questions should ask how strongly teachers agree with the following statements on a scale from (1) Strongly Disagree to (6) Strongly Agree:

- ➢ The leadership at my school supports hands-on science teaching
- ➢ Other teachers at my school support hands-on science teaching
- ➢ I have the support I need to lead my students in hands-on science teaching
- ➢ I have people I can go to if I have questions about hands-on science teaching

In addition, two questions should ask about the teacher's perception of continued support from PEN using the same agreement scale:

- ➢ PEN has provided a community for me focused on hands-on science teaching
- ➢ I can get help from PEN if I have questions about hands-on science teaching this year

Measures of teacher science identity and science teaching self-efficacy. A recent scoping review (Zai et al., 2024) noted that there are few validated survey measures of science teacher identity; most studies of science teacher identity have been qualitative (see Rushton & Reiss, 2021). One aspect of a science teacher identity is a science identity (Zai et al., 2024). Therefore, PEN can measure this aspect of science teacher identity with an instrument designed for student/individual science identity. Science identity frameworks are diverse, but many include similarities: (1) epistemic beliefs, (2) interest in science (3) recognition (see Chen & Wei, 2022; Guo et al., 2022). Self-beliefs, such as self-efficacy, are often included within these identity frameworks. We include self-efficacy as a separate motivational construct below with teacher-specific measures.

Epistemic beliefs. One widely-used epistemic beliefs scale that has been used with populations from elementary students through pre-service teachers is the Scientific Epistemic Beliefs (SEB) Questionnaire (Conley et al., 2004). In Conley et al. (2004), statements were presented with options on a five-point Likert scale from (1) Strongly Disagree to (5) Strongly Agree. To align with other metrics in the study and to avoid a central point, we suggest extending the scale to (6) Strongly Agree. The items are as follows:

Source

- ➢ Everybody has to believe what scientists say.
- ➢ In science, you have to believe what the science books say about stuff.
- ➢ Whatever the teacher says in science class is true.
- ➢ If you read something in a science book, you can be sure it's true.
- ➢ Only scientists know for sure what is true in science.

Certainty

- ➢ All questions in science have one right answer.
- ➢ The most important part of doing science is coming up with the right answer.

- ➢ Scientists pretty much know everything about science; there is not much more to know.
- ➢ Scientific knowledge is always true.
- ➢ Once scientists have a result from an experiment, that is the only answer.
- ➢ Scientists always agree about what is true in science.

Development

- ➢ Some ideas in science today are different from what scientists used to think.
- ➢ The ideas in science books sometimes change.
- ➢ There are some questions that even scientists cannot answer.
- ➢ Ideas in science sometimes change.
- ➢ New discoveries can change what scientists think is true.
- ➢ Sometimes scientists change their minds about what is true in science.

Justification

- ➢ Ideas about science experiments come from being curious and thinking about how things work.
- ➢ In science, there can be more than one way for scientists to test their ideas.
- ➢ One important part of science is doing experiments to come up with new ideas about how things work.
- ➢ It is good to try experiments more than once to make sure of your findings.
- ➢ Good ideas in science can come from anybody, not just from scientists.
- ➢ A good way to know if something is true is to do an experiment.
- ➢ Good answers are based on evidence from many different experiments.
- ➢ Ideas in science can come from your own questions and experiments.
- ➢ It is good to have an idea before you start an experiment.

Interest in science. To align results of this study with those from the international community, we recommend using items from the PISA science attitudes questionnaire that focused on interest and enjoyment in science (OECD, 2017). These questions were asked with respect to specific science topics and on a four-point agreement scale; we recommend asking them broadly about science, with the same six-point scale described earlier. We have also modified the first question in the teacher version below to focus on "teaching science"---the student version of the question notes "learning science." We have left the final question as reflecting learning to recognize that teachers continue to learn:

- ➢ I generally have fun when I am teaching science.
- ➢ I like reading about science.
- ➢ I am happy working on science problems.
- ➢ I enjoy acquiring new knowledge in science.
- ➢ I am interested in learning about science.

Recognition. For science recognition, we suggest a modified version of the Recognition scale within Chen and Wei (2022). The questions should be asked on the six-point Likert agreement scale. The original scale is seen in the student section, below. The suggested teacher items are:

- ➢ I think of myself as a science person.
- ➢ My peers recognize me as a science person.
- ➢ My school leaders recognize me as a science person.
- ➢ My family and friends recognize me as a science person.

Science teaching self-efficacy. There are a number of instruments to measure science teaching self-efficacy. One of the most popular is the *Science Teaching Efficacy Beliefs Instrument* (STEBI, Riggs & Enochs, 1989), which continues to be used today (e.g., Deehan & MacDonald, 2024). The pre-service teacher version of the STEBI has been used previously in Ghana (Coffie & Doe, 2019). Within the STEBI, eight items are focused on Personal Science Teaching Efficacy and eight items are focused on Science Teaching Outcome Expectancy. Given the STEBI's age, recent revisions have been undertaken to make the instrument shorter and clearer. If PEN would like to use a scale similar to the STEBI, we recommend the revised instrument based on the STEBI, the T-STEM Science Instrument (Unfried et al., 2022). These items are below and are presented as statements to agree with on a Likert-type scale:

T-STEM Personal Science Teaching Efficacy:

- ➢ I am continually improving my science teaching practice
- ➢ I know the steps necessary to teach science effectively
- ➢ I am confident that I can explain to students why science experiments work
- ➢ I am confident that I can teach science effectively
- ➢ I wonder if I have the necessary skills to teach science
- ➢ I understand science concepts well enough to be effective in teaching science
- ➢ Given a choice, I would invite a colleague to evaluate my science teaching
- ➢ I am confident that I can answer students' science questions
- ➢ When a student has difficulty understanding a science concept, I am confident that I know how to help the student understand it better
- ➢ When teaching science, I am confident enough to welcome student questions
- ➢ I know what to do to increase student interest in science

T-STEM Science Teaching Outcome Expectancy:

- ➢ When a student does better than usual in science, it is often because the teacher exerted a little extra effort
- ➢ When a student's learning in science is greater than expected, it is most often due to their teacher having found a more effective teaching approach

- When a low-achieving child progresses more than expected in science, it is usually due to extra attention given by the teacher
- If parents comment that their child is showing more interest in science at school, it is probably due to the performance of the child's teacher
- The inadequacy of a student's science background can be overcome by good teaching
- The teacher is generally responsible for students' learning in science
- If students' learning in science is less than expected, it is most likely due to ineffective science teaching
- Students' learning in science is directly related to their teacher's effectiveness in science teaching
- Minimal student learning in science can generally be attributed to their teachers

A number of researchers have criticized the STEBI family of instruments for not adequately capturing teachers' self-efficacy for inquiry science practices and therefore the Teaching Science as Inquiry (TSI) survey was developed by Smolleck et al. (2006). Although the TSI may capture science practices more aligned with PEN's theory of change, the instrument is 69 items long, making it impractical for administration. A version of the instrument is openly available within Smolleck's (2004) PhD thesis at: etda.libraries.psu.edu/files/final_submissions/3676

In order to capture teacher self-efficacy for PEN-specific practices, we recommend that the the following questions aligned with Bandura's (1977) theory of self-efficacy and with PEN's theory of change are asked using a six-point Likert agreement scale as in the prior measures:

- I can successfully match learning goals to hands-on activities for my science class
- I can identify the necessary materials for hands-on activities for my science class
- I can acquire the necessary materials for hands-on activities for my science class
- I can successfully explain science concepts through using hands-on activities
- I can engage my science students in learning with hands-on activities
- I can create my own hands-on activities to match new learning goals as they arise

To complement these questions, we advise PEN choose from the above science self-efficacy scales. The TSI's Personal Science Teaching Efficacy scale might be the best complement without creating too high a burden from a long survey.

Measures of teacher continued implementation. To measure teachers' continued implementation of hands-on activities, we suggest a two-pronged approach of surveys and observations. As noted in the sampling plan and procedures, above, all formerly-participating teachers can be provided with the survey. Suggested questions include:

- Do you still use the PEN Manual with sample hands-on activities? [yes/no]
- If yes…How often do you refer to the PEN Manual? [once a year or less, once a term, once a month, a few times a month, once a week, multiple times a week]

➢ How often do you implement hands-on activities from the PEN manual with your science class? [once a year or less, once a term, once a month, a few times a month, once a week, multiple times a week]
➢ How often do you implement any hands-on activities with your science class? [once a year or less, once a term, once a month, a few times a month, once a week, multiple times a week]

For the smaller observation sample, we recommend that PEN visit each teacher in the volunteer observation sample two times to observe when the teacher has indicated they will be implementing a hands-on lesson. During this time, PEN observers should use an observation tool developed based on critical components identified. This can be a *Fidelity Checklist* that is structured in a way to ensure that the activities are being conducted as intended *(See figure 2 for an example of a simple fidelity checklist).* Another tool that could be used is the *Fidelity Rating Scale*, which rates the degree to which each component of the intervention was carried out as intended. Ratings could be numerical or descriptive, such as “Fully Implemented” “Partially Implemented” and “Not Implemented.” For example, the scale could assess the level of teacher-student interaction or the extent of hands-on engagement during the activity. During these observation visits, PEN may want to employ Inter-Rater Reliability Checks, where they have more than one observer independently record the same teaching session, to check for consistency and reliability. These ratings are then compared, and discrepancies are noted and resolved to improve the overall reliability of the assessment.

Figure 2
An example of a Fidelity Checklist that can be adapted/modified to fit PEN’s need.

Component	**Yes/No**	**Notes**
Materials Used		
All required materials		
Proper use of materials		
Instructional Methods		
Clear instructions given		
Hands-on engagement		
Student Engagement		

Active participation		
Collaboration		
Adherence to Protocol		
Followed activity steps		

Measures of PEN interaction. These measures of interaction will depend on the availability of data on teacher engagement with PEN, but can include:

- Metrics from the WhatsApp group, including number of logins, number of posts or likes, time spent on the app, distribution of logins (e.g., how many months during the year was the teacher active).
- PEN-initiated contact and response, including how often PEN reached out to the teacher, how often they were able to contact the teacher via phone, and how often the teacher responded.
- Teacher-initiated contact, including how often the teacher reached out to PEN staff for assistance after the initial training.

All of these metrics may need to be adjusted to account for different amounts of potential time. For example, two points of contact in five years vs. in one.

Student Measures

Junior high school science engagement in hands-on activities. A few simple questions can be asked to capture whether students engaged in hands-on activities and/or were students of teachers who participated in PEN's training:

- What grade are you in now?
- What years did you attend junior high school?
- Where did you attend junior high school?
- The next questions ask about your use of hands-on activities in science class. By hands-on activities, we mean using physical items or tools to experiment with or represent a science concept.
    - When you were in junior high school, how often did you participate in hands-on activities in your science class? [never, once a year or less, once a term, once a month, a few times a month, once a week, multiple times a week]
    - How big a part of your junior high school science experience was hands-on activities? [not at all a part, a very small part, a small part, a mid-sized part, a large part, a very large part]

- ➢ How much did your junior high school teacher(s) use hands-on activities in science class? [never, once a year or less, once a term, once a month, a few times a month, once a week, multiple times a week]
- ➢ Provide an example of a hands-on activity you did in science class when you were in junior high school. [open ended]

Science self-efficacy. According to Bandura (1977), self-efficacy captures an individual's belief that they can execute the behavior necessary to produce a desired outcome. With respect to science for high school students, desired outcomes are often course grades and exam scores. We suggest PEN ask students who complete the survey the following questions:

- ➢ How confident are you that you can get Grade 1 in your science course this year? [not at all confident, slightly confident, somewhat confident, moderately confident, very confident, extremely confident]
- ➢ How confident are you that you can get better than a Grade 9 in your science course this year? [not at all confident, slightly confident, somewhat confident, moderately confident, very confident, extremely confident]
- ➢ How confident are you that you can get a grade A in the WASSCE? [not at all confident, slightly confident, somewhat confident, moderately confident, very confident, extremely confident]
- ➢ How confident are you that you can get at least a grade C6 in the WASSCE? [not at all confident, slightly confident, somewhat confident, moderately confident, very confident, extremely confident]

In addition, PEN may want to ask more general science self-efficacy and self-concept questions similar to those asked in other studies framed with SEVT (Eccles & Wigfield):

- ➢ How well do you expect to do in science this year? [not at all well, slightly well, somewhat well, moderately well, very well, extremely well]
- ➢ How good would you be at learning something new in science? [not at all good, slightly good, somewhat good, moderately good, very good, extremely good]
- ➢ If you could list all people in your class from the worst to the best in science, where would you place yourself? [the very bottom, near the bottom, in the middle, near the top, the very top]
- ➢ Compared to your other subjects, how much better at science are you? [much worst, a little worst, about the same, a little better, much better]

Self-efficacy is fed by individuals' prior successes with similar tasks, seeing similar peers accomplish the tasks, verbal persuasion from respected others, and physiological signs that the task is not too difficult (Bandura, 1977). PEN may also wish to ask students about the sources of their science self-efficacy using the *Sources of Science Self-Efficacy* scale (Britner & Pajares, 2006). The following are questions from that measure, in the order they were presented in the survey. Questions were asked using an agreement scale, (1) TOTALLY Disagree, (2) Mostly Disagree, (3) Somewhat Disagree, (4) Somewhat Agree, (5) Mostly Agree, (6) TOTALLY Agree. The specific source of self-efficacy is identified in parentheses after each item:

- I make excellent grades on science tests. (Mastery experience)
- Seeing adults do well in science pushes me to do better. (Vicarious experience)
- My science teachers have told me that I am good at learning science. (Social persuasion)
- Just being in science class makes me feel stressed and nervous. (Phys/Affective)
- I have always been successful with science. (Mastery experience)
- Many of the adults I know have jobs that involve science. (Vicarious experience)
- People have told me that I have a talent for science. (Social persuasion)
- Doing science work takes all of my energy. (Phys/Affective)
- I start to feel stressed-out as soon as I begin my science work. (Phys/Affective)
- Adults in my family have told me what a good science student I am. (Social persuasion)
- Even when I study very hard, I do poorly in science. (Mastery experience, Reverse scored)
- Seeing kids do better than me in science pushes me to do better. (Vicarious experience)
- I have been praised for my ability in science. (Social persuasion)
- My mind goes blank and I am unable to think clearly when doing science work. (Phys/Affective)
- Other students have told me that I'm good at learning science. (Social persuasion)
- People I admire are good at science. (Vicarious experience)
- I got good grades in science on my last report card. (Mastery experience)
- I do well on science assignments. (Mastery experience)
- The people I want to be like are mostly people who are involved in science. (Vicarious experience)
- My classmates like to work with me in science because they think I'm good at it. (Social persuasion)
- I get depressed when I think about learning science. (Phys/Affective)
- I compete with myself in science. (Vicarious experience)
- My whole body becomes tense when I have to do science. (Phys/Affective)
- I do well on even the most difficult science assignments. (Mastery experience)

Science identity. The Recognition scale within Chen and Wei (2022) can be used to measure students' science identity. The questions should be asked on the six-point Likert agreement scale:

- I think of myself as a science person.
- My classmates recognize me as a science person.
- My science teachers recognize me as a science person.
- My family and friends recognize me as a science person.

Science values. Science values can be represented by subjective task values with SEVT (Eccles & Wigfield, 2020). We suggest measuring interesting, utility, attainment, and cost values with the following items that are adapted from various studies of SEVT, including Gaspard et al. (2017) and Flake et al. (2015):

Interest Values:

- ➢ To you, how much fun is science? [six-point scale anchored with (1) Not at all fun and (6) Very fun]
- ➢ To you, how interesting is science? [six-point scale anchored with (1) Not at all interesting and (6) Very interesting]

Utility Values:

- ➢ How useful is science for your future career? [six-point scale anchored with (1) Not at all useful and (6) Very useful]
- ➢ How useful is science for your future coursework? [six-point scale anchored with (1) Not at all useful and (6) Very useful]
- ➢ How useful is science for your everyday life? [six-point scale anchored with (1) Not at all useful and (6) Very useful]

Attainment Values:

- ➢ How important is it to you to do well in science? [six-point scale anchored with (1) Not at all important and (6) Very important]
- ➢ How much do you care about what you learn in science? [six-point scale anchored with (1) Very little and (6) Very much]

Cost Values:

- ➢ How much do you have to give up to do well in science? [six-point scale anchored with (1) Very little and (6) Very much]
- ➢ How often do you feel negative emotions about science? [six-point scale anchored with (1) Never and (6) All the time]
- ➢ How much effort do you have to put in to do well in science? [six-point scale anchored with (1) Very little and (6) A lot]

Science intentions and enrollment. As PEN's theory of change and logic model note the ultimate outcomes of students engaging with hands-on activities in science class is enrollment and completion of science-focused university degrees and entry into the science workforce, we suggest asking a series of questions aligned with these outcomes:

- ➢ What is your ultimate career goal? [open ended]
- ➢ How likely is it that your career will involve science? [not at all likely, slightly likely, somewhat likely, moderately likely, very likely, extremely likely]
- ➢ Do you plan on attending university? [yes, no, maybe]

- ➢ How likely is it that you will attend university? [not at all likely, slightly likely, somewhat likely, moderately likely, very likely, extremely likely]
- ➢ What do you think your university major will be? [open ended]
- ➢ How likely is it that your university major will involve science? [not at all likely, slightly likely, somewhat likely, moderately likely, very likely, extremely likely]

Demographics. Given PEN's prior work investigates gender differences in the association of PEN activities and student outcomes (e.g., Beem, 2020), PEN should collect gender information on student survey respondents. In addition, differences across regions/districts within Ghana may influence results, so information on specific regions/districts should be collected. Finally, students are likely to be more similar to those in their specific class or school, so this information should also be collected.

Analysis

RQ1. Does student-reported hands-on learning relate to STEM motivation and outcomes among senior secondary students?

To answer this question, student responses regarding science self-efficacy, values, identity, and intentions can be used as outcomes. Each construct can be regressed separately on student use of hands-on activities, controlling for dummy variables for student year in secondary school, gender, and region. Depending on results of preliminary analyses described below, multilevel models may need to be used, with students nested within senior high classes and/or within schools. An example single-level regression is represented below:

Y[science identity] = B0[intercept] + B1X1[Hands-On Learning] +B2X2[Girl] + B3X3[Region2] + B4X4[Region3] +B5X5[Region4] + B6X6[Senior] +error

For a multilevel regression, a random intercept for class or school should be estimated.

In examining the results of these regressions, the coefficient of interest will be the hands-on learning independent variable. This coefficient will represent how students' reports of participating in hands-on learning in junior high school relate to their current science motivation and intentions. For example, in interpreting the results of this regression, one might say: for every point more on this hands-on learning measure, the students report [the value of B1] points higher on the science identity measure (or whichever measure is used as the outcome). This is what is called the "unstandardized coefficient," because it is in the scale of the specific measures. For example, if the value of B1 is .6, then one would say, for every point higher a student reports on the hands-on learning measure, they would report .6 points more on the science identity measure.

To be able to compare results across models and between this study and others, standardized beta coefficients should be calculated. For single-level regressions (not multilevel models), most statistics programs include a "beta" option that will automatically provide the beta coefficient. When using multilevel regressions, this should be hand calculated and can be done using the formula B(SDX)/SDY, where the unstandardized coefficient, B is multiplied by the standard deviation of X (in this case, the hands-on learning measure) and the resulting product is divided by the standard deviation of Y, the outcome (in this case, the science identity measure). These standardized beta coefficients are interpreted in standard deviation units and can therefore be compared across models and even across studies. For example, if our B1 is .6 and our standard deviation of X is .75 and of Y is .90, then the standardized beta coefficient would be .6(.75)/.90 or .5 standard deviations. In this case, one would interpret the standardized coefficient by saying for every one *standard deviation* increase in reported hands-on learning, there is half a *standard deviation* increase in reported science identity.

If PEN collects data on the multiple outcomes suggested above, there may be other more complex analyses that could also be helpful. For example, treating the different motivation variables as completely independent constructs may not adequately capture the profile of a student who is "motivated for science." Person-centered analyses, such as latent profile analysis (e.g., Hong et al., 2020) may better represent the data. In this way, profiles of students with different levels of each of the motivational constructs can be created and profile membership can be predicted from reported hands-on science experiences. Instructing on this method is outside the scope of this report, but PEN may wish to investigate the possibility with a data analyst.

Before running regressions or other analyses, a number of preliminary data steps should be undertaken.

(1) Descriptive statistics should be calculated for all items to ensure that no items are out of range and to examine patterns of missing data.
(2) Scales should be created for the motivation constructs (i.e., science self-efficacy, science identity, science task value components, science intentions) and the hands-on activities use items by taking the mean average of each set of questions. Internal consistency of constructs should be calculated with a metric, such as Cronbach's alpha (Cortina, 1993).
(3) The science intentions outcome items should be examined and combined to form a reasonable outcome. Depending on student answers, one possibility might be to code open-ended career goals for science involvement; another is to use the students' rating of likelihood that their career and/or major would involve science, either as individual variables or as a combined scale (mean averaged).
(4) ICCs should be examined for each outcome. Unconditional (null) multilevel models can be estimated by regressing the outcome on no predictors but including a random

intercept for class or school. Higher ICCs (above even .05) indicate that running single-level models may present biased results (see Julian, 2001) and therefore multilevel models should be used.

RQ2. Does enrollment in a classroom in which teachers had received PEN training relate to STEM motivation and outcomes among senior secondary students?

Similar methods as in RQ1 can be used to prepare the data, but regressions for this question will include a dummy variable for whether the student was enrolled in a school with a PEN-trained teacher in place of the student use of hands-on activities scale. This dummy variable will provide the coefficient of interest and would be interpreted as the presence of participation in PEN programs. For example, a similar regression to the one noted above for hands-on learning would be:

Y[science identity] = B0[intercept] + B1X1[0/1 if PEN] +B2X2[Girl] + B3X3[Region2] + B4X4[Region3] +B5X5[Region4] + B6X6[Senior] +error

If B1 in this regression equaled .6, then one would interpret it as: Compared to students who did not have PEN-trained teachers, those with PEN-trained teachers reported science identities .6 points higher.

To standardize these coefficients as betas, the formula is only B(1)/SDY, because the variable is a 0/1 variable indicating the presence or absence of experience with PEN-trained teachers. In this case, given the same .9 standard deviation of Y (science identity), the beta coefficient would be .6/.9 or .667. This coefficient could be compared to beta coefficients from other interventions that also have results for motivation outcomes and can be situated relative to these interventions.

RQ3. To what extent are teachers' use of hands-on activities maintained after the PEN training?

This is largely a descriptive research question and PEN can examine means, standard deviations, and ranges of teachers' reported continued implementation and motivation (science teaching self-efficacy, science identity). For example, PEN can calculate percentages in each of the bins for the following questions or can treat the latter three questions as continuous variables and calculate a mean for each and for a scale combining all three.

- Do you still use the PEN Manual with sample hands-on activities? [yes/no]
- If yes…How often do you refer to the PEN Manual?
- How often do you implement hands-on activities from the PEN manual with your science class?
- How often do you implement any hands-on activities with your science class?

RQ 4. To what extent do teachers engage with PEN after their initial training?

As with RQ3, this is largely a descriptive research question. PEN can create two scales from the measures of community engagement, one from the two PEN community engagement questions and one from the four other community engagement questions. Means can be examined to determine what support teachers report.

RQ 5. To what extent does this engagement and other community support predict teachers' continued implementation and positive motivations?

The descriptive analyses from RQ3 and RQ4 can be extended to understand factors that are associated with teacher-reported continued implementation by choosing one or more of the implementation variables and regressing these on either the measure of PEN community support or the measure of teacher community support. This would be a single-level regression.

For example, PEN may combine the latter three implementation questions into a "hands-on learning" scale variable and regress this variable on a scale created from the two PEN community support questions and on a scale created from the four other community support questions. PEN will likely want to include teacher demographic factors, such as years teaching and school factors, such as region.

Y[implementation] = B0[intercept] + B1[PEN Support] + B2[Other Support] + B3X3[Region2] + B4X4[Region3] +B5X5[Region4] + B6X6[Years Teaching] +error

Coefficients can be interpreted as described above for RQ1.

Similarly, teacher motivation variables (science identity, science teaching self-efficacy) can be regressed on support variables to see if these variables also predict teacher motivation.

RQ6. To what extent does positive motivation predict continued implementation?

Finally, it may be that teachers with high science teaching self-efficacy or high science identity are more likely to implement hands-on lessons. To examine these relationships, PEN can regress the implementation variables as in RQ5 on motivation variables, for example:

Y[implementation] = B0[intercept] + B1[Science Identity] + B2[Science Teaching Self-Efficacy] + B3X3[Region2] + B4X4[Region3] +B5X5[Region4] + B6X6[Years Teaching] +error

However, these results should be interpreted with caution given the surveys are being collected at the same time. Although it may be reasonable to conclude that teachers with higher science and science teaching motivation are more likely to implement hands-on learning, the reverse may also be true—that implementation of hands-on learning improves science and science teaching motivation.

Deliverable 2: Organization Growth Recommendations

Introduction

This deliverable outlines actionable recommendations to support PEN's organizational growth, with a focus on enhancing program fidelity, scaling outcomes, and aligning with long-term strategic objectives. These recommendations build on insights from the tracer study and PEN's existing evidence-based practices to foster sustainable development and scalability.

Justification of Need

PEN's cascading model is heavily reliant on the Training Of Trainers Approach (over 3,000 Trainers) for scaling the teachers' training (8,763 teachers trained) and students reach (2M+ students reached) within and beyond Ghana's border. PEN also recently began to diversify its approach, adding on the Exemplar Teachers Program, with increased partnerships, and having multiple training formats (Mini training, standard training, Stages 1-6, In-person/Online/Blended). As the Organization begins to grow, scale, and ensure the validity of the program impact, the need to improve program fidelity becomes imperative.

Improving Program Fidelity

Improving Program Fidelity for Educational Initiatives refers to the systematic approaches aimed at ensuring that educational interventions are implemented as designed to achieve evidence-based outcomes. This concept has gained prominence as educational stakeholders recognize that high fidelity in program delivery is crucial for maximizing effectiveness and ensuring reliable results and is a necessary part of evaluating interventions (Century et al 2010; Stains & Vickrey 2017; Gale, *et al.* 2020) . Variability in the execution of educational programs can lead to inconsistent outcomes, highlighting the importance of fidelity as a key factor influencing the success of interventions. Program fidelity refers to the degree to which educational initiatives adhere to their prescribed components, dosage, and quality of implementation, all of which are essential for evaluating effectiveness and optimizing student engagement and success in STEM fields. (Lemire et al., 2022; Gale, et al., 2020; Process Street, n.d). Gould et al. (2014) posit that in order to achieve intended program outcomes, program implementers need to understand the essential and indispensable components that define a program's success. Noting that outcomes assessment alone, however, is not sufficient to build a rigorous evidence base for intervention practices but a systematic study of *fidelity of program implementation* (*FOI*) is needed to provide a more robust understanding of the core components of these interventions, their potential to improve specified teacher and student outcomes, and our ability to implement these programs consistently and effectively over time and in diverse school settings

Conceptual Framework

The theoretical foundation for improving program fidelity in educational initiatives, particularly those focused on STEM teaching, relies heavily on established frameworks, such as the Fidelity of Implementation (FOI) model. This model provides structured guidelines for evaluating the adherence to program components, dosage delivered, and the quality of implementation, which are essential for understanding program fidelity (Carroll et al., 2007; Century et al., 2010, Lemire et al., 2022).

Researchers and program developers must identify the primary components of an intervention and explain how they are expected to generate the intended results in order to conduct an investigation into FOI. (Abry et al., 2014; Capin et al, 2018). Subsequently, it is essential to develop reliable and valid measures of FOI, as these criteria can then be used in further studies to evaluate experimentally if differences in how key components are implemented and are consistently linked to certain outcomes across replication trials. (Century et al 2010)

By breaking down programs into their fundamental elements, evaluators can identify the core components critical for successful implementation, ensuring a comprehensive approach to measuring effectiveness (Process Street, n.d).

Applying structured frameworks for examining innovation implementation can enhance fidelity in educational settings. (Barker et al. 2014). By clearly defining what constitutes effective implementation and aligning assessment criteria with these components, practitioners can ensure adherence to best practices and evidence-based strategies.

There are four commonly agreed-upon dimensions of fidelity from our review of literature, which are *Adherence, Dosage, Quality, and Participants Responsiveness.*

- Adherence. This refers to the extent to which the intervention is implemented as it was designed.
- Dosage. This refers to the amount or frequency of the intervention delivery.
- Quality. This refers to the degree to which the intervention was implemented with competence and skill.
- Participants Responsiveness. Also known as treatment receipt, refers to the extent to which participants engage with and respond to the intervention.

For PEN, considering the four basic dimensions of program fidelity questions to be raised could include

- **Adherence**: Were all the components of the intervention delivered as intended? Were the intervention materials used as prescribed? This relates both to the Training Program and the Teachers’ Classroom practice. It depends on the level at which this is being considered. Meaning when PENs TOTs/Teachers facilitate the training was it delivered as intended by the trainers at that level, and at the Teachers level, were the hands on strategies facilitated as intended.

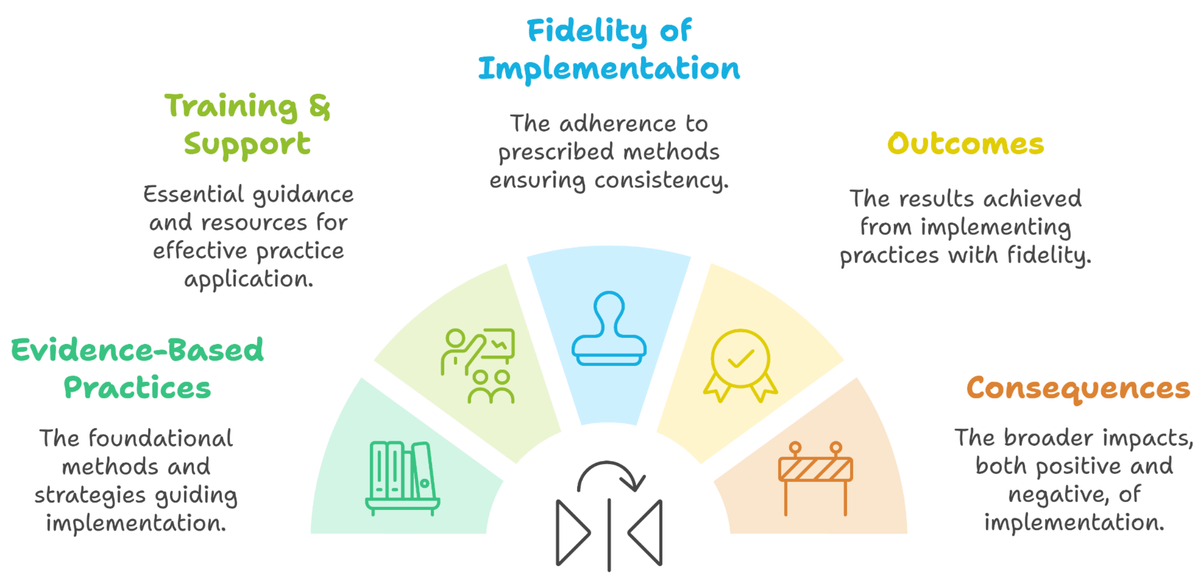
- **Dosage**: Was the intervention delivered for the recommended amount of time? Were participants exposed to a sufficient amount of the intervention to achieve the intended outcome?
- **Quality**: Was the intervention delivered with fidelity to the instructional strategies, techniques, and principles of the intervention? Were the teachers able to adjust the intervention in response to students' needs?
- **Participants Responsiveness**: Did the participants actively participate in intervention activities? Were participants able to demonstrate understanding and application of the intervention content?

Treatment fidelity data (descriptive and statistical) are critical to interpreting and generalizing outcomes of intervention research (Capin et al., 2018; Lakin & Rambo-Hernandez, 2019). Programs developed without a clear instructional framework often leave teachers with substantial decision-making responsibilities, resulting in inconsistent effectiveness across different educational contexts (Carroll et al., 2007). Understanding these limitations is crucial for refining existing programs to better support educators in achieving intended outcomes.

In order to achieve the intended program outcomes, program implementers need to understand the essential and indispensable components that define a program's success. Some of these components as extracted from the literature reviewed here are represented in the figure below.

Figure 3

Fidelity of Implementation



Notes on Fidelity of Implementation

Evidence-Based Practices (EBPs)

- Selection: Choosing interventions with proven effectiveness
- Adaptation: Modifying EBPs to fit the context without losing core elements.
- (*To learn more about implementing an EBP with fidelity, view the IRIS Module*: Evidence-Based Practices (Part 2): Implementing a Practice or Program with Fidelity)

Training and Support:

- Providing robust training and ongoing support for educators is essential in promoting high-fidelity implementation.
- Ensure that everyone is receiving the same training and support, with the aim that the delivery of the intervention is as uniform as possible. Such strategies include the provision of manuals, guidelines, training, and monitoring and feedback for those delivering the intervention.

Fidelity of Implementation Components:

- Adherence, quality of delivery, participant responsiveness, and program differentiation. (A*s discussed above)*

Outcomes:

- Immediate Outcomes: Observable changes shortly after implementation (e.g., student engagement, knowledge gain).
- Intermediate Outcomes: Changes that occur after some time (e.g., skill development, behavior change)
- Long-term Outcomes: Sustained impacts (e.g., improved academic performance, career readiness).

Consequences:

- Positive Consequences: Enhanced learning, better retention rates, increased interest in STEM.
- Negative Consequences: If fidelity is low, potential for ineffective practices and wasted resources.

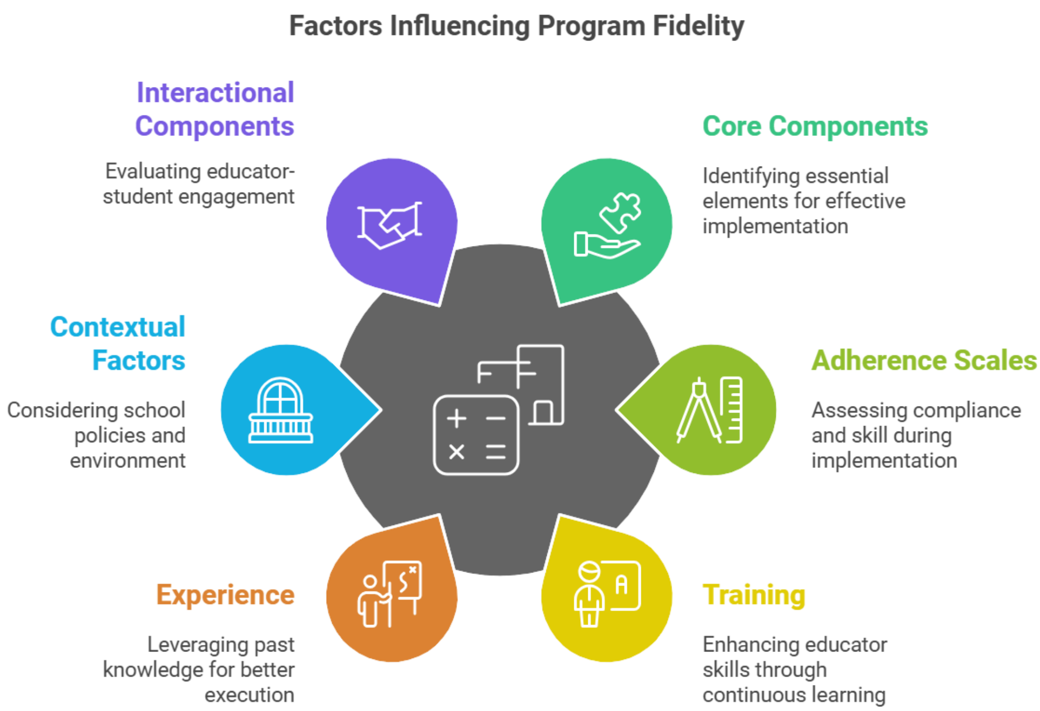
Factors influencing Program Fidelity

Studies that evaluate the implementation process (Abry et al., 2014; Baker et al., 2014; Capin et al., 2018; Gould et al., 2016; Lakin & Rambo-Hernandez, 2019) have identified several factors that may influence or moderate the degree of fidelity with which an intervention is implemented. The figure below highlights some of these factors that could explain the presence or absence of intervention effects and how far an intervention actually affects outcomes.

Understanding the factors influencing program fidelity is essential for evaluating the effectiveness of PEN's STEM teaching interventions and ensuring that desired educational outcomes are achieved.

Figure 4

Factors Influencing Program Fidelity



Measuring Fidelity of Implementation

Measuring program fidelity encompasses the several dimensions that must be systematically measured to provide a comprehensive understanding of how interventions are executed and their resultant impact on outcomes (Century et al., 2010). To effectively measure these dimensions, various methods can be employed, which includes:

Observation-based assessments involve direct observation of program activities, real-time monitoring, to ensure that they are being implemented as intended.

Self-report measures rely on participants or implementers reporting their adherence to program guidelines by utilizing self-assessment tools.

Surveys can be used to collect feedback from stakeholders about the fidelity of program delivery.

Document analysis entails reviewing program-related documents and examining written records, such as lesson plans or training materials, to assess fidelity.

Implementation logs serve as detailed records of program implementation, aiding in the evaluation process by adhering to predefined criteria and standards.

For further reading on measuring fidelity of implementation:
https://www.process.st/how-to/measure-fidelity-of-implementation/

Additional Recommendations

Digital Resource Libraries

The creation of digital resource libraries to serve as centralized repositories of teaching aids and materials, including lesson plans, activity guides, and video demonstrations. By leveraging digital resources, PEN can reduce the costs associated with producing manuals. This cost-effectiveness allows for the allocation of funds to other critical areas, such as more professional development offers.

Training participants and master trainers can have a digital repository which could be an online folder or platform where they have continuous access to relevant materials to reference back to, as at when needed. Access to these materials may have diverse access restrictions based on the different training stages. Also a repository where relevant resources can be shared or uploaded by the PEN teachers' community can also be created. The idea here is that beyond the training period, when teachers know that they can always access these resources from anywhere and at any time, there is a higher tendency of use. Secondly, materials and resources beyond those of PENs but are relevant to assisting teachers understand the theories and instructional practices underpinning PENs interventions can also be shared by PEN or the community of teachers. One of the greatest benefits of digital resource libraries is their accessibility. This convenience is especially valuable for educators in remote or underserved areas who may have limited access to traditional teaching aids.

Community Forums

Professional Learning Communities (PLCs): Foster a collaborative learning environment and a community of practice for the trained teachers by setting up online forums or social media groups where these teachers can ask questions, share experiences, successes, resources, strategies and support each other. Encourage peer mentoring, group activities, collaborative projects to build a supportive network.

Ongoing Support: Ongoing support after teacher training is crucial. It is recommended that PEN have dedicated staff whose responsibility is to offer regular check-ins or follow-up sessions to help teachers implement what they've learned and to address any challenges they face. This can also be a means to regularly assess the impact of the training on teaching practices and student learning outcomes. This includes encouraging teachers to engage in self-reflection and self-assessment, conducting regular performance evaluations and providing constructive feedback can help teachers identify areas for improvement.

Recommendation for the Community of Practice (CoP)

This is recommended as "not a formal space," but a space where Master Trainers/Trainees share, learn, ask and connect.

- **Monthly/bi-monthly/quarterly exchange?** A live session of 1 hour where 3-4 members of the different cohorts will conduct activities, demonstrations, discussions or make presentations for 10-15 minutes each.

- **How can they participate:** By being a presenter/facilitator, an attendee, or both.

- **Possible formats for sessions:** Presentation, demonstration, hosting a conversation around a theme/question, participatory activity, anecdotes or story that we can learn from.

This platform should allow beneficiary teachers to

- Know others who are taking the PEN course
- Develop a **network** by interacting and connecting with teachers of the PEN program from across geographies
- Allows a shared context for you to communicate, **exchange ideas** and share information
- Allows asynchronous sharing of ideas
- Enable **dialogue** amongst practitioners who want to hone their hands-on skills
- Stimulate **learning** by serving as a vehicle for communication and cross-learning
- **Capture** and diffuse existing knowledge and practices
- Promote **collaborative problem-solving** and encourage the free flow of ideas and strengthen purposeful action

Note: *This approach of peer learning is inspired by Open Space Technology (OST) Design. OST is participant-driven and less organizer-convener-driven*

"The goal of an Open Space Technology meeting is to create time and space for people to engage deeply and creatively around issues of concern to them. " Ideally, an Open Space Technology facilitator is neither seen nor heard, but his or her presence is "felt." - ***Chris Corrigan***

A more detailed approach to building community to engage teachers is outlined in the last section of this report.

Deliverable 3: Future Research Options

Introduction

This deliverable explores future research directions to deepen understanding of PEN's program outcomes and inform ongoing improvement. By leveraging robust research methodologies, these recommendations aim to generate actionable insights that align with PEN's mission to foster hands-on STEM education and contribute to broader educational research.

Research Types for Educational Interventions

Foundational Research, **Early-Stage** or **Exploratory Research** contributes to core knowledge in education.

Efficacy, Effectiveness, and **Scale-up Research** (*Impact Research*) contributes to evidence of impact, generating reliable estimates of the ability of a fully developed intervention or strategy to achieve its intended outcomes.

Based on PEN's stage of organizational growth, it is believed that they already have a developed intervention and therefore are more interested in impact research.

Purpose of Studies that Assess the Impact of Education Interventions and Strategies

The purpose of impact studies is to generate reliable estimates of the ability of a fully developed intervention or strategy to achieve its intended outcomes. For an impact study to be warranted, the theory of action must be well established and the components of the intervention or strategy well specified. For all impact studies, descriptive and exploratory analyses should be sufficiently elaborated to determine the extent to which the findings support the underlying theory of action. The types of impact studies include:

Efficacy research allows for testing of a strategy or intervention under "ideal" circumstances, including a higher level of support or developer involvement than would be the case under normal circumstances.

Effectiveness research examines effectiveness of a strategy or intervention under circumstances that would typically prevail in the target context, without substantial developer involvement in implementation or evaluation.

Scale-up research examines effectiveness in a wide range of populations, contexts, and circumstances, without substantial developer involvement in implementation or evaluation.

As PEN continues to seek funding for their programs through impact studies, these guidelines will be instrumental in increasing their success in preparing grant proposals. Additionally, these guidelines will aid the PEN team in developing a clearer understanding of what different stages of education research should address and what outcomes they should produce. Which, in turn, will support more informed decision-making based on the level of evidence provided.

Efficacy research evaluates how well an intervention works under ideal and controlled conditions. This means the study is conducted in a highly controlled environment, often with strict protocols and a selected group of participants. The goal is to determine if the intervention can produce the desired effect when everything is perfect. Efficacy research is best used in the early stages of testing a new intervention, to establish its potential effectiveness under optimal conditions. The PEN has conducted this type of research in Ghana (Beem 2020 study) and would be ideal for entry in new locations like Liberia or establishing efficacy of new models.

Effectiveness research, on the other hand, assesses how well an intervention works in real-world settings. This type of research takes place in more naturalistic environments, with a broader and more diverse population, and less control over variables. For PEN the aim is to understand how the intervention performs in everyday classroom practice, without support from PEN trainers. Effectiveness research is best used after efficacy has been established, to see how the intervention works under typical school conditions.

Scale-up research is best used when there's strong evidence from efficacy and effectiveness research that an intervention works well in controlled and real-world settings, respectively. The goal here for PEN is to ensure that these interventions can be effectively replicated in different contexts and can be implemented widely and sustainably.

Figure 5

Guidelines for Justification and Evidence of the Studies

Type of Research	Purpose	Implementation Support	Policy and/or Practical Significance	Justification	Theoretical and Empirical Basis
Efficacy Research	To determine whether an intervention or strategy can improve outcomes under what are sometimes called "ideal" conditions	May include more implementation support or more highly trained personnel than would be expected under routine practice,	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.	The proposal should justify the choice to examine the impact of the intervention under ideal implementation conditions with a well defined sample, rather than under routine practice conditions.	Efficacy Research should be justified by one or more of the following: (1) empirical evidence of the promise of the intervention from a well designed and implemented pilot study. (2) empirical evidence from at least one well-designed and implemented Early-Stage or Exploratory Research study supporting all the critical links in the intervention's theory of action; (3) evidence the intervention is widely used even though it has not been adequately evaluated to determine its efficacy; or (4) if the intent is to replicate an evaluation of an intervention with a different population, evidence of favorable impacts from a previous well-designed and implemented efficacy study and justification for studying the intervention with the new target population.
Effectiveness Research	To estimate the impacts of an intervention or strategy when implemented	Study should be carried out with no more developer involvement than what would be expected under	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address;	The proposal should justify the choice to examine the impact of the intervention under ideal implementation	Effectiveness Research should be justified by strong empirical evidence of the efficacy of the intervention, as demonstrated by statistically significant and substantively important estimates of impact, from one study that includes multiple sites or settings, or two studies

	under conditions of routine practice	typical implementation.	(2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions	conditions with a well defined sample, rather than under routine practice conditions.	that each include one site or setting, all of which meet the guidelines for evidence to be produced by Impact Research (Table 4) or evidence that the intervention is widely used even though it has not been adequately evaluated for efficacy.
Scale-up Research	To estimate the impacts of an intervention or strategy under conditions of routine practice and across a broad spectrum of populations and settings.	Should be carried out with no more developer involvement than what would be expected under typical implementation.	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.	The proposal should justify the choice to examine the impact of the intervention under typical implementation conditions with a broad sample, rather than under ideal implementation conditions with a well-defined sample or under routine practice conditions with a relevant typical sample.	Scale-up Research should be justified by compelling evidence of the effectiveness of the intervention, as demonstrated by statistically significant and substantively important impact estimates from one study that includes multiple sites or settings, or two studies that include one site or setting, all of which meet the guidelines for evidence to be produced by Impact Research (Table 4). In addition, there should be no overriding evidence demonstrating a negative impact of the intervention

Project Outcomes: Efficacy, Effectiveness, and Scale-up reporting should include detailed descriptions of the study goals, design and implementation, data collection and quality, and analysis and findings.

Study reports should document implementation of both the intervention and the counterfactual condition in sufficient detail for readers to judge applicability of the study findings. When possible, these factors should be related descriptively to the impact findings. Study reports should discuss implications of the findings for the theory of action and, where warranted, make suggestions for adjusting the theory of action to reflect the study findings. If a favorable impact is found, the project should identify the organizational supports, tools, and procedures that were key features of the intervention implementation. If no evidence of a favorable impact is found, the project should examine possible reasons (e.g., weaknesses in the implementation, evidence that raises questions about particular aspects of the logic model).

Research Plan: Efficacy, Effectiveness, and Scale-up research plan should identify and justify (1) the study design used to estimate causal impact of the intervention on the outcomes of interest; (2) the key outcomes of interest for the impact study and the minimum size impact of the intervention that would have policy or practical relevance; (3) the study setting(s) and target population(s); (4) the sample, including the power it provides for detecting an impact; (5) the data collection plan, including information about procedures and measures, including evidence on and strategies for ensuring reliability and validity, and plans for collecting data on program implementation, comparison group practices, and study context; and (6) the analysis and reporting plan.

For Impact Research (as opposed to Design and Development Research), quasi-experimental designs, such as matched comparison groups or regression discontinuity designs, are acceptable only when there is direct compelling evidence demonstrating the implausibility of common threats to internal validity. These might include selection bias in the case of matched comparison groups. Ideally, the study sample size and allocation to condition should be such that the minimum true impact detectable size with 80 percent power and a 95 percent confidence interval is no larger than the minimum relevant size impact for policy or practice. If that is not the case, the proposal should provide a rationale for conducting the study despite its not meeting this standard.

Primary outcome measures should include student outcomes sensitive to the performance change the intervention is intended to bring about. student outcomes not strictly aligned with the intervention, and student outcomes of practical interest to educators and policymakers.

The project should measure the strength and qualities of implementation (sometimes referred to as “fidelity of implementation”) to address whether the intervention’s impact estimates may be linked to how it was implemented.

The project should measure comparison group practices and/or conditions to support a clear characterization of the contrast between the intervention and comparison condition.

The analysis plan should specify analytic models that reflect the sample design and maximize the likelihood of obtaining unbiased, efficient estimates of average impacts and the confidence intervals around those impacts. The analysis plan should describe additional analyses conducted to explore variability in the intervention's impacts and possible implications for the theory of change.

Sources : [Common Guidelines for Education Research and Development]

Specific Recommendations for PEN's Future Studies

Randomized Control Trials

The strongest causal conclusions for efficacy or effectiveness research can be made from *randomized control trials* (RCTs). In this type of research design, teachers or schools would be randomly assigned to receive PEN training (or not). Without random assignment, even if PEN were to measure many covariates that might influence teacher success (e.g., location, years training, teacher science experience), it would be impossible to ensure that all teachers in the two groups (PEN/not PEN) were equal. When these groups are unequal, the worry is that an effect is attributed to PEN that instead is due to something else (e.g., teacher background, motivation, resources). When teachers are randomly assigned to treatment and control conditions, all these background variables (both measured and unmeasured) are also randomly assigned across conditions. With a large enough sample size, the two conditions can be assumed to be equal.

What is a large enough sample size? A power analysis is necessary to help determine the appropriate size. For single-level studies (e.g., with teacher-level outcomes), a program such as G*Power can be helpful in estimating the necessary number of teachers in conditions. G*Power is a free, open source software available at this link. When student outcomes are used, the nesting of the study (students within teachers) complicates the power analysis, and more sophisticated methods should be used, such as Optimal Design power analysis software. In order to estimate the necessary sample size, PEN must first determine a number of parameters:

What is the expected effect size? This can be determined by referencing other research in similar locations or by leveraging results from the tracer study. For example, the tracer study may show that the average association between self-reported engagement with PEN programs and later student science motivation is .3 standard deviations (a .3 effect size). This would present a good starting point for calculating power.

What is the typical class size? To calculate power accurately, the average class size for teachers' classes (both participating and control) is needed.

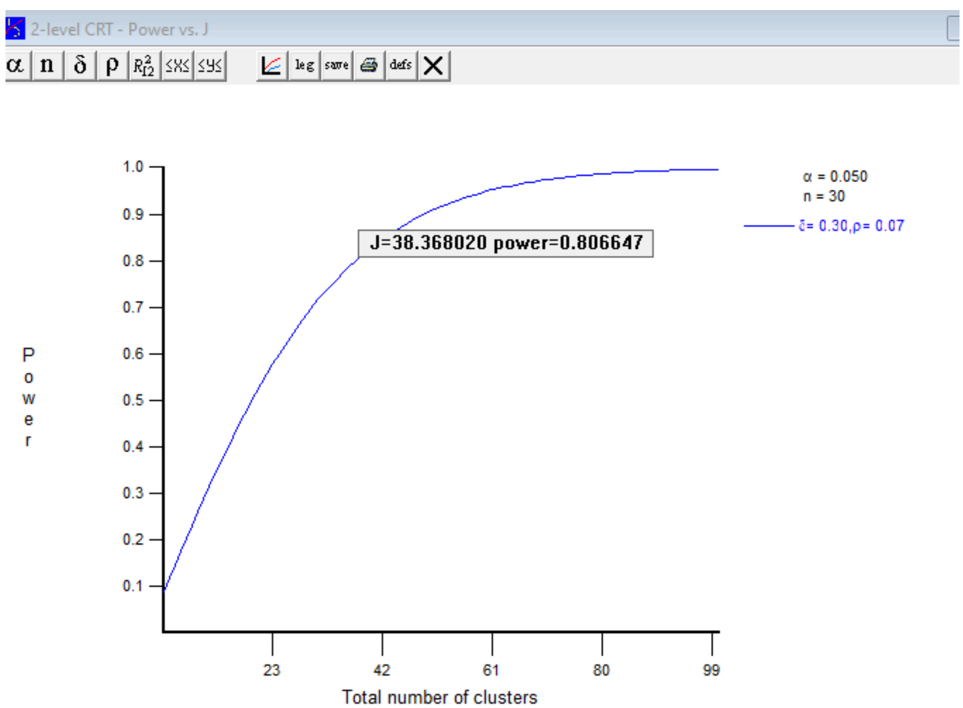
What is the typical ICC? The intraclass correlation represents how similar students are in the same classes on the outcome variable of interest. Prior research in similar contexts can help determine the appropriate number here. For example, prior research may show that among classes of junior secondary students in Ghana, an ICC of .07 is typical for science self-efficacy.

Other parameters? For other parameters, such as power and alpha level, field norms can be used here. Typically, 80% power is specified and an alpha of .05.

Using Optimal Design to estimate power for a cluster randomized trial (treatment assigned at level 2, the teacher level) with student-level outcomes and the sample numbers provided above, the software reports that 38 teachers are needed for 80% power, assuming an average class size of 30 students.

Figure 6

Screenshot of Optimal Design



In recruiting teachers for an RCT, PEN should recruit teachers as similar to each other as possible and randomly assign them to treatment and control through a method like a random number chart. It will be important to not allow teachers to switch conditions in order to maintain fidelity to the assignment and take advantage of the benefits to internal validity from the RCT. To increase teacher buy-in and participation, PEN can offer the teachers selected for the control an opportunity to receive the training the following year.

After assignment to conditions is completed, PEN should administer pre-tests and pre-surveys to the teachers and students in the study, both treatment and control. Then, PEN should complete the training of the treatment teachers, monitor fidelity as noted in Deliverable 2, and finally, administer post-tests and post-surveys to all teachers and students in the study. If PEN uses a delayed treatment model and offers control teachers PEN training the following year, it is advisable to continue to measure teachers and students in the study with an additional round of pre- and post-tests/surveys. This way, PEN can determine if the effect of the program two years after training is the same as one year and if the effect in the delayed treatment group is the same as the original treatment group. This information strengthens PEN's claims regarding the effectiveness of their program.

Data can be analyzed with models similar to those in the Tracer Study RQ2 analysis. The variable as to whether the student had a PEN-trained teacher would be replaced with whether the student's teacher was *assigned* to PEN training. This is called an Intent to Treat analysis, as it only considers each teacher's assignment and not whether the teacher actually completed all training or implemented hands-on lessons.

Quasi-Experimental Designs

If PEN is unable to carry out an RCT, there are adjustments to selection procedures for quasi-experimental designs that can be implemented to strengthen claims of PEN's efficacy or effectiveness. With methods, such as the one used in Babb and Stockero (2020), where the coach assigns teachers to treatment and control, there is high concern for selection bias (e.g., coaches might select "better" teachers to the treatment or teachers they particularly want to work with). To the extent possible, decisions about which schools are in treatment and which are in control should be as far removed from coaches and teachers as possible. Some other possibilities include:

Waitlist control: PEN can have teachers sign up for training and assign any after a certain number of spots to the waitlist or delayed treatment list. This ensures that all teachers who sign up are motivated and interested in PEN. However, there may be relationships between teaching effectiveness and earlier sign-up.

Matched controls: PEN can recruit all teachers who are interested and assign teachers to treatment and control based on similarity (each treatment teacher should have a control teacher with similar background, training, and students). One way PEN can implement matched controls is to have all teachers' students take the pre-test/pre-surveys and match based on these variables.

In quasi-experimental designs, it will be especially important for PEN to administer pre-tests and pre-surveys to ensure that the groups are, overall, equivalent on these variables and to control for prior scores/responses, especially if they are not. It will also be important to collect information on the teachers, including their teaching background, training, location, prior science motivation and attitudes toward and knowledge about hands-on learning.

Longitudinal Studies

As PEN considers implementation of methods to ensure fidelity and maintenance of hands-on task implementation as recommended in Deliverable 3, these methods can also serve as structure for longitudinal studies. In particular, Digital Resource Libraries and Community Forums provide a means to contact engaged teachers after they have completed PEN training and may not otherwise be in contact with PEN staff. Methods like experience sampling (Hektner et al., 2007) can be embedded in such platforms. In experience sampling, short surveys are displayed to participants (like teachers accessing a digital resource library) to ask about their experience of using the service or their current state. For example, teachers who access the resource library might, on some visits, be asked “Why did you visit the resource library today?” or “Are you visiting the library to plan for a hands-on lesson? How soon is the lesson? [today, this week, next week, within the month, next month, later than next month].” Asking these sorts of questions when teachers access resources can help PEN understand the value of these resources and collect data on how frequently teachers plan to use hands-on activities.

Experience sampling can also be used as part of surveys sent to teachers. For example, once a month PEN could send a survey that asks teachers, “Have you taught a hands-on lesson this month?” and/or “How many hands-on lessons have you taught this month?” The value of these brief surveys as opposed to a survey at the end of a year or longer period is that teachers are more likely to accurately recall more recent experiences and therefore the quality of the data gathered will be higher.

With regards to longitudinal surveys of students, one barrier is that students frequently change their cell phones or other contact information. PEN should obtain each student’s email address so that they can continue to follow them after they have left junior secondary school. To encourage student participation, PEN could offer students who share their contact information and/or engage with PEN surveys entry into small drawings, such as for gift cards or technology prizes.

PEN can also secure information from teachers each year as to where their existing students are going to attend senior secondary and can target those schools for follow-up surveys. Similarly, once these students are in their final year of senior secondary, school rosters can be used to secure information on the students’ next academic move. Maintaining these databases of students and school locations will require investment in database management software and personnel, but may pay large dividends in being able to answer more definitively whether gains made immediately after PEN programming are maintained through senior secondary and university and into careers. The measures and analytic methods recommended for the Tracer Study can be modified and applied to longitudinal studies, although additional questions about current science involvement will likely be helpful to add.

Deliverable 4: Contextualizing & Communicating Effects

Contextualizing the Tracer Study Results

Effective communication of program outcomes is essential for ensuring stakeholder engagement and understanding. PEN seeks to contextualize and share the outcomes of its hands-on STEM education programs with diverse audiences, including educators, funders, and government officials. By aligning its findings with audience priorities, PEN can highlight its unique contributions to STEM education and its ability to address critical gaps in teaching and learning.

To achieve this, PEN must go beyond activities reporting, which emphasizes implementation details, to adopt an outcome-oriented reporting approach that demonstrates meaningful, measurable results. This shift positions PEN to clearly articulate not just what was done but the difference these actions have made for learners, teachers, and the education system. The following sections outline strategies for presenting these results effectively to diverse stakeholders.

Positioning Activity and Outcome Reporting

Communicating program outcomes effectively requires distinguishing between activities reporting and impact reporting. While activities reporting highlights what was done, such as quantitative metrics and implementation details, impact reporting delves into the outcomes and evidence of meaningful change resulting from those activities. For PEN, this distinction is crucial to demonstrate how its programs translate into tangible benefits for learners, teachers, and the broader education system.

For example:

- **Activities Reporting**: “PEN trained 100 teachers in hands-on STEM methodologies.”
- **Outcome Reporting**: “Students taught by these teachers demonstrated increased engagement, as evidenced by classroom observations and surveys.”

By focusing on outcome reporting, PEN can showcase measurable improvements for teachers and students, aligning its efforts with broader educational goals and stakeholder priorities.

Models for Good Communication of Impact

Once standardized results have been calculated and contextualized, the next step is to present these findings using models that resonate with diverse stakeholders, ensuring clarity and engagement.

Data Visualization

Graphs and charts are effective tools for conveying trends and comparisons:

- Bar Charts: Illustrate pre- and post-intervention outcomes.
- Heatmaps: Highlight geographic variation in program reach.
- Infographics: Present complex data in simplified, visually appealing formats for non-technical audiences.
- Scatterplots and Line Graphs: Highlight relationships between variables, such as increased hands-on learning and higher science identity, with annotations for standardized effect sizes.

Case Studies and Testimonials

Qualitative data adds depth to quantitative findings. Stories from teachers, students, and school leaders illustrate the tangible outcomes of PEN's programs, complementing statistical analysis. By connecting these stories to evidence of outcomes—such as improved science identity or increased hands-on teaching adoption—PEN can move beyond activities reporting ("What was done?") to impact reporting ("What difference was made?").

Comparative Analysis

Positioning PEN's findings within the broader landscape of educational initiatives in sub-Saharan Africa demonstrates its potential value. For instance, highlighting how PEN's cost-efficient model could achieve results comparable to or exceeding those of other programs strengthens its position among stakeholders.

Depth of Evidence

Outcome reporting emphasizes both observable and sustainable changes that programs like PEN's aim to achieve. For PEN, a tracer study could explore and highlight:

- **Observable Trends**: Potential measurable improvements in student engagement, science identity, and motivation associated with hands-on learning.
- **Sustainable Practices**: Long-term shifts in teacher practices, such as the adoption and consistent use of hands-on methodologies, which the study may identify as prevalent trends among PEN-trained teachers.
- **Alignment with Systemic Goals**: Evidence of how PEN's methods complement Ghana's evolving curriculum reforms, offering a cost-effective and scalable solution for enhancing STEM education.

By investigating these areas, the tracer study seeks to build a narrative around how PEN's programs may support systemic educational improvements. While the findings cannot establish causality, they can offer valuable insights into the relationships between PEN's training programs and observed outcomes in classrooms and among students.

Designing a One-Pager for Communicating Tracer Study Results

A well-designed one-pager can serve as a concise, visually appealing tool for summarizing PEN's findings and engaging stakeholders. This format is particularly effective for funders and policymakers who need key information at a glance.

Key Components of the One-Pager

1. **Headline or Phrase.** A compelling headline that encapsulates the findings. Examples include:
    - "*From Classroom to Careers: How PEN's Training Shapes STEM Aspirations*"
    - "*Sustaining STEM: The Long-Term Influence of Hands-On Teaching*"
    - "*Inspiring Science Identity: Evidence from PEN's Hands-On Approach*"
    - "*Catalyzing Change in STEM Education: Insights from PEN's Tracer Study*"
    - "*Innovating STEM Classrooms: Findings from PEN's Teacher Training Program*"
2. **2–3 Key Statistics.** Highlight meaningful outcomes that address tracer study questions. Examples include:
    - **Student Outcomes**: *"Students in classrooms with PEN-trained teachers reported a 0.6 SD increase in science identity, indicating a significant boost in STEM motivation."*
    - **Teacher Retention of Practices**: *"85% of teachers reported sustained use of hands-on activities one year after PEN training."*
    - **Community Support Predicting Implementation**: *"Teachers with high engagement in PEN's alumni network were 70% more likely to continue implementing hands-on activities in their classrooms."*
3. **Quote or Testimonial.** Include qualitative feedback to complement the quantitative data:
    - *"PEN's training has completely changed how I approach science lessons. My students are more engaged than ever before."—Junior High Science Teacher, Ghana.*
    - *"Hands-on learning is making science real for our students—they're starting to see themselves as future engineers and scientists."—School Administrator, Kumasi.*
4. **Visuals.** Use visuals to summarize key findings.

    - **Bar Chart**: Show pre- and post-training outcomes, such as the increase in science identity or teacher confidence.
    - **Infographic**: Map teacher retention rates for hands-on activities across regions.
    - **Comparison Chart**: Highlight differences in outcomes between students taught by PEN-trained teachers and those in non-PEN classrooms.
5. **Takeaway Message.** Summarize the broader implications of PEN's work in one sentence. For example:
    - *"PEN's training equips teachers to deliver practical, impactful STEM education, preparing students for future success.*

Suggested Examples Addressing Tracer Study Questions

- **Does student-reported hands-on learning in junior high school relate to STEM motivation and outcomes among senior high students?**
    - Statistic: *"Students reporting hands-on learning in junior high school demonstrated a 0.4 SD increase in STEM intentions by senior high school."*
    - Visual: Scatterplot showing the relationship between hands-on learning frequency and STEM intentions.
- **Does enrollment in a classroom in which teachers had received PEN training relate to STEM motivation and outcomes among senior high students?**
    - Statistic: *"Enrollment in PEN-trained classrooms increased STEM career aspirations by 30% compared to non-PEN classrooms."*
    - Visual: Heatmap indicating STEM outcomes by region for PEN-trained classrooms.
- **To what extent are teachers' use of hands-on activities maintained after the PEN training?**
    - Statistic: *"75% of teachers continued using PEN's hands-on STEM activities in their classrooms after one year."*
    - Visual: Bar chart illustrating hands-on activity retention rates over time.
- **To what extent do teachers engage with PEN after their initial training?**
    - Statistic: *"80% of PEN-trained teachers participated in alumni network discussions at least once per quarter."*
    - Visual: Infographic depicting alumni engagement activities, such as webinars and resource-sharing.
- **To what extent does this engagement and other community support predict teachers' continued implementation and positive motivations?**
    - Statistic: *"Teachers actively engaged in PEN's alumni network were twice as likely to report confidence in teaching STEM hands-on."*
    - Visual: Pie chart showing the breakdown of teacher motivations by alumni engagement levels.
- **To what extent does positive motivation predict continued implementation?**

    - Statistic: *"Teachers with high motivation were 1.5 times more likely to implement hands-on learning consistently across all STEM topics."*
    - Visual: Line graph comparing teacher motivation scores and implementation rates.

Suggested Design Principles

- **Simplicity**: Avoid clutter; focus on a clean, professional layout.
- **Brand Consistency**: Use PEN's colors, fonts, and logo to reinforce its identity.
- **Accessibility**: Ensure visuals are easy to interpret for non-technical audiences.

By crafting a one-pager with these elements, PEN can create an engaging, digestible summary of its tracer study results that appeals to a broad range of stakeholders.

Framework for Presenting Standardized Metrics

To ensure consistency and comparability, PEN should employ standardized metrics, such as effect sizes, to contextualize its findings within broader educational research:

- **Standardized Beta Coefficients:** Beta coefficients interpret relationships in terms of standard deviation units, enabling comparisons across models and studies. These metrics allow PEN to demonstrate measurable outcomes in an accessible, quantifiable format (see Deliverable 1 for formulaic details and examples).

  Example: "For every one standard deviation increase in reported hands-on learning, there is a corresponding half a standard deviation increase in reported science identity."

- **Dummy Variables:** To evaluate the presence or absence of PEN's programs, dummy variables simplify analysis.

  Example: "Students in classrooms with PEN-trained teachers reported science identity approximately two-thirds of a standard deviation higher than their peers without PEN-trained teachers."

- **Data Preparation and Quality Control:** Before conducting regressions or calculating effect sizes, PEN should:
    - Calculate descriptive statistics to check for data accuracy and identify any missing data patterns.

- Combine related survey questions into a single score to measure key concepts, such as teacher confidence or hands-on teaching practices. This ensures that each measure is reliable and accurately represents the intended idea.
- Examine response patterns within groups, such as students in the same class. If responses are highly similar within groups, apply methods that account for this clustering to ensure accurate and unbiased results.

By employing these models and metrics, PEN can effectively communicate its program outcomes, fostering greater understanding and support for its initiatives.

Review of PEN's External Communication Materials

Overview

PEN provided their external communication materials for review, requesting feedback to enhance clarity, coherence, and alignment with strategic goals. This section summarizes key strengths, identifies areas for improvement, and offers actionable recommendations.

Strengths

PEN's communication materials demonstrate several commendable qualities, including:

- **Consistency in Messaging and Design**: The materials maintain a uniform narrative, format, and color scheme, reinforcing PEN's brand identity.
- **Clear Articulation of the Training-of-Trainer Model**: This model is effectively positioned as a cost-efficient approach, aligning with PEN's overarching narrative of scalability and affordability.

Opportunities for Improvement

Several areas within the materials could be refined to strengthen their impact:

- **Clarify the Link Between Curriculum and Rote Learning**: While the materials highlight the prevalence of rote learning and reference the Ghanaian national curriculum's emphasis on practical activities, they could more explicitly state that curriculum reform alone does not address the issue. Emphasizing the need for teacher and school leader training to implement these changes effectively would underscore the importance of PEN's program.
- **Integrate Short- and Long-Term Impact Narratives**: Highlighting both immediate outcomes (e.g., student engagement) and long-term effects (e.g., career success in STEM fields) would provide a more comprehensive view of the program's value.
- **Strengthen the Connection to Economic Growth**: The materials mention STEM as a driver of GDP but could better link this to Ghana's lack of service industries and the

potential for PEN's approach to contribute to long-term economic development. However, this connection should be concise to avoid overcomplication.

- **Clarify Strategic Partnerships**: The materials could provide more detail about partnerships with NACCA and GAST. Explaining how these relationships enable scaling and implementation would strengthen the narrative.

Recommended Changes to Existing Slides

To enhance the effectiveness of PEN's slide deck:

- **Use Authentic Imagery**: Replace AI-generated photos of students with images of Ghanaian students who have participated in PEN's programs. If such photos are unavailable, consider addressing this within the pitch, stating, "We currently use AI-generated visuals but hope to feature actual program participants as we continue to grow."
- **Streamline Text**: The slides detailing student impact contain excessive text and could benefit from concise summaries supported by visuals. Simplifying these slides would improve readability and retention.
- **Highlight Complementary Nature of the Program**: Emphasize that PEN's training integrates seamlessly with existing teacher training initiatives, requiring minimal additional resources and making it a cost-effective solution.
- **Incorporate Key Events**: Include details from the Liberia work and STEM Symposium at the World Bank into the master deck to showcase PEN's global engagement.

Additional Feedback and Recommendations

Beyond slide-specific adjustments, broader feedback was collected to enhance PEN's strategic positioning.

Develop One-Pagers

Creating concise, program-specific one-pagers tailored for potential funders can enhance follow-up communication. These one-pagers should focus on key elements such as the problem, PEN's solution, unique value proposition, program impact, and scalability. By streamlining the main deck, this approach ensures it remains succinct and focused on high-level messaging.

Address Text-Heavy Infographics

Simplifying text-dense infographics will make them more visually appealing and accessible to diverse audiences. Prioritize clarity by highlighting key points, reducing unnecessary details, and incorporating visuals to summarize complex information.

Clarify Teacher Capacity-Building vs. Ease of Teaching

It remains unclear whether PEN's approach primarily builds teacher capacity or facilitates ease of teaching. Clarifying this distinction is critical for positioning PEN's program as a scalable, skill-building initiative that equips educators with durable and transferable teaching methodologies. This refinement could also serve as a foundation for expanding the program to other topics beyond STEM.

During the review process, several questions and comments arose that could further enhance PEN's communication materials and strategic positioning:

- **Clarify "Our Approach" in the Peace Corps PEN 2024 Presentation Deck**: The explanation of "our approach" in the deck is unclear, particularly the reference to "money coming back." A clearer articulation of this concept would improve understanding and alignment with the broader narrative.
- **Simplify PEN Infographics**: While the infographics effectively use PEN's branding, they are text-heavy. Streamlining these visuals would make them more accessible and engaging for diverse audiences.
- **Highlight the Role of School Leaders**: The example from one school illustrates the importance of involving school leaders in program implementation. Expanding on this could underscore the systemic impact of PEN's model.
- **Examine Impact Through Three Lenses**: PEN's materials could benefit from framing its impact across three dimensions:
    1. **Immediate Impact**: Current improvements in student learning outcomes.
    2. **Long-Term Impact**: The lasting influence on students' life trajectories and career choices.
    3. **Ecosystem Building**: Contributions to creating a broader culture of STEM engagement and innovation.

Communicating with the Government

Scaling PEN's programs to a national level requires strategic engagement with government stakeholders. This section outlines a broad framework PEN can consider for scaling its work, drawing from examples in Global School Leader's *How to Scale with Government: A Toolkit for Education Organizations in the Global South*. These insights aim to spark a deeper conversation about institutionalizing PEN's work within government systems.

Key Opportunities for PEN

PEN's unique position and approach create several opportunities for effective government engagement:

- **Effective and Scalable Solution**: PEN's training model is easy to implement and adaptable for broader adoption.
- **Cost-Effectiveness**: The program's affordability aligns well with budgetary constraints faced by governments.

- **Alignment with Curriculum Reforms**: PEN complements Ghana's evolving curriculum guidelines by emphasizing hands-on teaching methods.
- **Political Timing**: Upcoming elections present a potential opening to align with new government priorities.

Framework for Scaling with Government

The process of scaling with governments can be broken into five key stages:

1. **Align on the Mindsets**[1]
   - Determine if working with governments aligns with PEN's vision and approach. While systemic shifts can create the greatest impact, this path requires long-term commitment and resilience.
   - Acknowledge that scaling is often non-linear, requiring PEN to explore multiple paths and remain flexible in responding to opportunities.

2. **Identify the Opportunity**[2]
   - **Map the System**: Identify key government offices, roles, and priorities. This mapping is especially critical during periods of political change, such as upcoming elections.
   - **Highlight Cost Efficiency**: Frame PEN's solution in terms of concrete budgetary benefits for governments. Demonstrating scalability with specific numbers can strengthen buy-in.
   - **Prepare for the Unexpected**: Opportunities cannot always be predicted. By creating system maps and public trackers, PEN can stay prepared to act when windows of opportunity open.
   - **Align Narratives**: Reframe PEN's messaging to align with government priorities and decision-makers' goals.

3. **Tap into Partners**[3]
   - **Build Coalitions**: Partner with organizations and individuals whose missions align with PEN's goals. Co-creating initiatives with allies fosters momentum and strengthens advocacy efforts.
   - **Expand Networks**: Leverage PEN's international supporter base and connections to build coalitions at national and global levels. For example, international funders and supporters can amplify PEN's credibility and influence.

4. **Do the Work**[4]

---

[1] See slides 6, 8, 10, and 15 of the toolkit for more information and examples
[2] See slides 13, 14, 19, 20, 23, 17, and 42 of the toolkit for more information and examples
[3] See slides 31, 32, 36, 37, 45, 46, 47, 48, 49, and 50 of the toolkit for more information and examples
[4] See slides 17, 18, 26, 35, 55 and 57 of the toolkit for more information and examples

- **Deliver Proof Points**: Demonstrate program effectiveness through data and case studies that resonate with government priorities. Identify unique data that only PEN can provide to highlight program success.
- **Engage in Experiential Advocacy**: Create opportunities for government officials to experience PEN's methods firsthand. In-person engagements can leave lasting impressions and build political goodwill.
- **Work Across Levels**: Engage with multiple layers of government simultaneously to ensure broad support and maximize opportunities.

5. **Establish the Work**[5]
   - **Secure Quick Wins**: Deliver short-term successes that governments can claim as their own. This creates incentives for officials to support and sustain PEN's programs.
   - **Embed the Program**: Identify the right place within the government system for PEN's initiatives to become permanently institutionalized. Focus on building political will and transferring credit to government stakeholders.

Engaging with Program Alumni

Engaging teachers from PEN's alumni network is a key focus area for fostering ongoing learning, collaboration, and professional development. By maintaining connections with alumni, PEN can not only share updates, news, and resources but also create a vibrant community where teachers reflect on their practices, share experiences, and solve challenges together. This sense of community empowers educators to continue learning and innovating, even if they are not actively participating in discussions, as they can still draw inspiration and insights from their peers' contributions.

Types of Online Communities

Professional development communities typically take two forms:
1. **Standalone Communities**: Interest-based groups where participants connect independently to share ideas, discuss challenges, and learn from each other. These groups can be hosted on platforms like WhatsApp or Facebook and supplemented with occasional virtual gatherings via Zoom or Google Meet to deepen engagement.
2. **Add-On Communities**: Extensions of existing programs that facilitate continuous learning, offer peer or coach support, and foster long-term connections among participants.

PEN's Online Community Approach

[5] See slides 28, 29, 40, and 41 of the toolkit for more information and examples

Based on pilot tests conducted by Global School Leaders (GSL), WhatsApp has proven highly effective for national and regional community-building, offering an accessible and user-friendly platform. For PEN, focusing on WhatsApp-based communities will provide a foundation for building robust alumni networks while aligning with educators' familiarity and access to this tool.

Framework for Building an Effective Online Community

PEN can create a thriving online alumni community by following these steps:

1. **Building the Foundations**
   Define the purpose and goals of the community. Key considerations include:
    - Why should the community exist, and how will it benefit participants?
    - What outcomes does PEN want to achieve—individual professional growth, collective learning, or both?
    - Is the community a fresh initiative, a continuation of a previous effort, or an extension of an existing program?

   Key Decisions:
    - **Target Audience**: Identify who will participate, their motivations, and potential barriers to engagement.
    - **Group Structure**: Define whether the community will be open or closed, its size, and the preferred mode of interaction (e.g., synchronous or asynchronous).
    - **Technology Platform**: Ensure the chosen platform, such as WhatsApp, is accessible, intuitive, and supported by participants' infrastructure.
    - **Time Frame**: Decide if the community is time-bound or ongoing. For open-ended communities, establish periodic themes or focus areas to sustain engagement.
    - **Facilitation**: Assign a dedicated facilitator to manage the community, communicate effectively, and ensure smooth operation.

2. **Planning**
   Once foundational decisions are made, PEN can design the community's structure and content.
    - **Community Guidelines**: Set clear expectations for participation, acceptable content, and consequences for violations (e.g., no political or religious content).
    - **Engagement Routines**: Establish predictable patterns, such as weekly innovation sharing (e.g., "Thursdays at 5 PM are for sharing your classroom innovation").
    - **Content Calendar**: Plan themes, learning cycles, and resources in advance (e.g., "In January, we will focus on strategies for hands-on STEM experiments"). Maintain a backup of content for contingencies.

- ○ **Resource Directory**: Create an easily accessible repository for shared materials, ensuring participants can quickly find relevant resources.
- ○ **Engagement Metrics**: Define metrics to track success, such as participation rates, message volume, and resource clicks.

Lay Out the Culture Plan:
Culture is the shared set of beliefs, assumptions, and values that influence participant behavior and interactions within the community. A strong culture fosters trust, collaboration, and ownership among members. PEN's culture plan should include:

- **Defining Values and Expectations**: Clearly articulate the shared goals, norms, and values that guide interactions (e.g., "We value respectful dialogue and collaboration").
- **Reinforcing Values Through Activities**: Incorporate routines, systems, and activities that emphasize the desired culture (e.g., weekly posts highlighting collaborative achievements or success stories).
- **Empowering Participants**: Create opportunities for members to take ownership of their learning and contribute to the community (e.g., assigning champions to lead discussions or share resources).

3. **Launching**
Launching the community effectively sets the stage for sustained participation and engagement. A successful launch requires careful preparation and execution to generate excitement and establish clear expectations. Key steps include:
    - **Recruit Participants**: Ensure participants are informed about the purpose and benefits of the community beforehand and have given their consent to join. Pre-launch communication is essential to set the tone and align expectations.
    - **Manage the Database**: Maintain an organized record of participants using tools such as Google Sheets or Google Contacts. Accurate tracking will help monitor invitations, responses, and membership.
    - **Create Buzz and Excitement**: Use promotional materials, teasers, or countdowns to build anticipation for the launch. This could include sharing highlights of what members can expect, such as expert-led sessions or resource-sharing opportunities.
    - **Onboard Participants**: Track invitations sent, accepted, and declined to manage membership effectively. Provide clear instructions for joining and outline the community's objectives and expectations during the onboarding process.
    - **Host a Dynamic Launch Event**: Begin with icebreaking activities and introductions to foster a sense of belonging. For example, encourage participants to share a recent experience using hands-on STEM tools (e.g., "Share one way you've used

STEM tools in the last month"). Use the event to establish guidelines, highlight the community's purpose, and drive initial engagement.

4. **Facilitating and moderating**

Ensuring the smooth operation of the community requires proactive facilitation and thoughtful moderation. Effective management fosters a positive environment, encourages participation, and addresses challenges as they arise. Key responsibilities include:

- **Reinforce Community Guidelines**: Regularly remind participants of the community's guidelines and enforce them consistently. If a member is removed for violating the rules, communicate the reason clearly to the group and reiterate the expectations.
- **Regulate Conversations**: Strike a balance between guiding discussions and allowing participants the freedom to express themselves. Facilitate open, respectful dialogue while addressing behaviors that may hinder constructive engagement. For example, ensure debates remain professional and do not devolve into personal criticism.
- **Respond to Emerging Needs**: Stay attuned to participant feedback and trends within the group. Regularly read messages to identify common themes, challenges, or requests. Act promptly to address concerns, such as implementing new rules for private messaging if participants report discomfort (e.g., "Participants should seek consent publicly before sending direct messages").
- **Encourage Engagement**: Actively catalyze discussions by posing questions, summarizing key points, and sharing fresh, relevant resources. For example, "What strategies have worked best for you in implementing STEM activities with limited resources?"
- **Provide Feedback and Recognition**: Show appreciation for meaningful contributions by tagging participants and acknowledging their input. Highlighting individual or group achievements reinforces engagement and motivates continued participation.

Guidelines for Effective Facilitation and Moderation

The success of an online community hinges on skilled facilitation and proactive moderation. Facilitators must balance planning with adaptability to create an environment that fosters engagement and trust. Below are practical strategies for effective community management:

- **Stay Organized**: A facilitator often juggles multiple tasks, some planned and others spontaneous. Keeping a structured approach ensures timely responses to inquiries and builds credibility with participants.
- **Understand the Context**: Recognize the diverse backgrounds and experiences of participants to tailor discussions and resources. Segmenting the community based on demographic data or emerging trends (e.g., gender, grade level, or geography) can help address specific needs effectively.

- **Catalyze Discussions**: Spark meaningful dialogue by asking open-ended questions, using visual elements, and summarizing key points. For example, ask, "What hands-on STEM activity has resonated most with your students?"
- **Share Relevant Resources**: Regularly provide content aligned with the group's interests and goals. A well-curated mix of resources encourages ongoing engagement.
- **Acknowledge Contributions**: Publicly recognize participants who add value to discussions. Tagging contributors or awarding digital badges can motivate others to engage actively.
- **Leverage Champions**: Identify and empower community champions to lead discussions or share insights, amplifying the facilitator's efforts.
- **Track Engagement Metrics**: Utilize raw data to analyze group conversations and interactions. Insights, trends, and progress can be tracked using the following tools for WhatsApp - (a) ChatAnalyzer, ChatVisualizer, and Chatlizer. Key metrics to measure include active participants, inactive participants, messages shared over a week, survey responses, and document or video views/downloads (e.g., via Bitly or Jumpshare). Additionally, consider tracking engagement through poll responses.

Optimizing Content and Engagement

To sustain interest and maximize participation, facilitators should consider best practices::

- **Use Positive Reinforcement**: Tagging specific participants to thank them for their contributions fosters a culture of appreciation. For example, acknowledging a teacher's innovative resource-sharing can inspire others to participate.
- **Engage with Visuals**: Posts incorporating graphics, thumbnails, or emojis (e.g., "📚 Resource Corner" or "🎥 Video Corner") tend to attract more attention. Pairing visuals with questions further boosts interaction.
- **Prioritize Relevant Resources**: Materials that address immediate needs, such as updates on curriculum reforms, often see higher engagement. For instance, during the COVID-19 pandemic, a storybook on child safety measures garnered more interest than reopening guidelines.
- **Focus on Peer-Curated Content**: Community-created resources, such as video tutorials or lesson plans, generate significant engagement and foster a sense of ownership.
- **Limit Resource Overload**: Sharing one resource per message ensures that all links receive attention. When multiple links are included, only the first tends to get substantial engagement.

Measure Community Success

To evaluate the effectiveness of the online community, PEN should adopt a multi-faceted approach to measuring engagement:

- **Track Conversations**: Assess the number and quality of discussions generated by shared resources.
- **Monitor Resource Interaction**: Use tools like Bitly to count clicks on links or track time spent on materials hosted online.
- **Analyze Participation Trends**: Compare the activity levels of participants over time to identify patterns or areas for improvement.

The Role of the Facilitator

Effective facilitation requires planning and daily community management by one facilitator (for consistency). We saw evidence that active facilitation increased the response and discussion levels. Facilitators need to:

- Create an engagement routine to build a familiar pattern for the participants. Example could be assigning a specific activity to specific work days (but it does not have to be done every day, in fact, 2-3 times a week in the M-W-F or Tu-Th pattern work best)
- Plan a content calendar to organize and schedule content to be shared. Different strategies like sequencing the resources or spiraling, or scaffolding can be used to plan based on the complexity level of the resources. Developing a bank of pre-written messages formatted (with emoticons and links) in one document can make executing the calendar easier.

Conclusion

By following these strategies, PEN can build a dynamic and engaged alumni network that supports professional growth, fosters collaboration, and amplifies the impact of its STEM education programs. A well-facilitated online community not only strengthens connections among educators but also positions PEN as a leader in innovative, scalable educational initiatives.

This vibrant network will serve as a powerful example of PEN's commitment to fostering sustainable change in STEM education, setting the stage for further collaboration and broader impact.

References

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2014). Using indices of fidelity to intervention core components to identify program active ingredients. American Journal of Evaluation, 36(3), 320–338. https://doi.org/10.1177/1098214014557009

Avraamidou, L. (2014). Studying science teacher identity: Current insights and future research directions. *Studies in Science Education, 50*(2), 145-179.

Babb, J. J., & Stockero, S. L. (2020). Impact of Practical Education Network on students in selected Ghanaian junior high school science classrooms. *African Journal of Research in Mathematics, Science and Technology Education, 24*(2), 216-228.

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

Barker, B., Nugent, G., & Grandgenett, N. (2014). Examining fidelity of program implementation in a STEM-oriented out-of-school setting. International Journal of Technology and Design Education, 24, 39-52. https://doi.org/10.1007/S10798-013-9245-9.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287-311.

Beem, H. (2020). Effect of hands-on science activities on Ghanaian student learning, attitudes, and career interest: A preliminary control study. *Global Journal of Transformative Education, 2,* 18-31.

Bendtsen, M., Forsman, L., & Björklund, M. (2021). Exploring empowering practices for teachers' sustainable continuing professional development. *Educational Research*, 64, 60 - 76. https://doi.org/10.1080/00131881.2021.2000338.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *43*(5), 485-499.

Capin, P., Walker, M.A., Vaughn, S. et al. (2018). Examining How Treatment Fidelity Is Supported, Measured, and Reported in K–3 Reading Intervention Research. Educ Psychol Rev 30, 885–919 https://doi.org/10.1007/s10648-017-9429-z

Carroll, C., Patterson, M., Wood, S. et al. (2007). A conceptual framework for implementation fidelity. Implementation Sci 2, 40 https://doi.org/10.1186/1748-5908-2-40

Century, J., Rudnick, M., & Freeman, C. (2010). A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge. American Journal of Evaluation, 31(2), 199-218. https://doi.org/10.1177/1098214010366173

Chen, S., & Wei, B. (2022). Development and validation of an instrument to measure high school students' science identity in science learning. *Research in Science Education, 52*(1), 111-126.

Chris Corrigan. Planning an Open Space Technology meeting https://www.chriscorrigan.com/parkinglot/planning-an-open-space-technology-meeting/?form=MG0AV3

Coffie, I. S., & Doe, N. G. (2019). Preservice teachers' self-efficacy in the teaching of science at basic schools in Ghana. *Science Teacher*, *10*(22), 101-106.

Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology, 29*(2), 186-204.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control?. Clinical psychology review, 18(1), 23-45.

Deehan, J., & MacDonald, A. (2024). Relations between Australian primary teachers' approaches to student choice and their reported science teaching efficacy beliefs and teaching practices. *International Journal of Science Education*, *46*(2), 181-203.

Dira-Smolleck, L. (2004). *The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry* (Doctoral dissertation, The Pennsylvania State University). etda.libraries.psu.edu/files/final_submissions/3676

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232-244.

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. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

Gale, J., Alemdar, M., Lingle, J. et al. (2020). Exploring critical components of an integrated STEM curriculum: an application of the innovation implementation framework. IJ STEM Ed 7, 5 https://doi.org/10.1186/s40594-020-0204-1

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology, 48*, 67-84.

Global School Leaders. https://www.globalschoolleaders.org/policytoolkit

Gould, L. F., Dariotis, J. K., Greenberg, M. T., & Mendelson, T. (2016). Assessing Fidelity of Implementation (FOI) for School-Based Mindfulness and Yoga Interventions: A Systematic Review. Mindfulness, 7(1), 5–33. https://doi.org/10.1007/s12671-015-0395-6

Gould, L. F., Mendelson, T., Dariotis, J. K., Ancona, M., Smith, A. S., Gonzalez, A. A., ... & Greenberg, M. T. (2014). Assessing fidelity of core components in a mindfulness and yoga intervention for urban youth: applying the CORE Process. New directions for youth development, 2014(142), 59-81.

Guo, X., Hao, X., Deng, W., Ji, X., Xiang, S., & Hu, W. (2022). The relationship between epistemological beliefs, reflective thinking, and science identity: a structural equation modeling analysis. *International Journal of STEM Education, 9*(1), 40.

Hampden-Thompson, G., & Bennett, J. (2013). Science teaching and learning activities and students' engagement in science. *International Journal of Science Education, 35*(8), 1325-1343.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). Experience sampling method: Measuring the quality of everyday life. Sage.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.

Holstermann, N., Grube, D., & Bögeholz, S. (2010). Hands-on activities and their influence on students' interest. *Research in Science Education, 40*, 743-757.

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology, 112*(7), 1409.

Institute of Education Sciences (Ed). (2013). Common guidelines for education research and development. https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf

Iversen, O., & Dindler, C. (2014). Sustaining participatory design initiatives. *CoDesign*, 10, 153 - 170. https://doi.org/10.1080/15710882.2014.963124

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8*(3), 325-352.

Lakin, J. M., & Rambo-Hernandez, K. (2019). Fidelity of Implementation: Understanding Why and When Programs Work. Gifted Child Today, 42(4), 205-214. https://doi.org/10.1177/1076217519862327

Lemire, C., Rousseau, M., & Dionne, C. (2022). A Comparison of Fidelity Implementation Frameworks Used in the Field of Early Intervention. American Journal of Evaluation, 2023, Vol. 44(2) 236-252. DOI: 10.1177/10982140211008978

Luehmann, A. L. (2007). Identity development as a lens to science teacher preparation. *Science Education, 91*(5), 822-839.

Madden, L., & Wiebe, E. N. (2013). Curriculum as experienced by students: How teacher identity shapes science notebook use. *Research in Science Education, 43*, 2567-2592.

Meesuk, P., Wongrugsa, A., & Wangkaewhiran, T. (2021). Sustainable Teacher Professional Development Through Professional Learning Community: PLC. *Journal of Teacher Education for Sustainability*, 23, 30 - 44. https://doi.org/10.2478/jtes-2021-0015.

OECD. (2017). Scaling procedures and construct validation of context questionnaire data. PISA 2015 technical report.

Pandey, U.C., & Kumar, C. (2019). Engagement with the Community and Sustainable Development. In: Leal Filho, W. (eds) Encyclopedia of Sustainability in Higher Education. Springer, Cham. https://doi.org/10.1007/978-3-319-63951-2_344-1

Process Street. How To Measure Fidelity Of Implementation - https://www.process.st/how-to/measure-fidelity-of-implementation/

Riggs, I. M., & Enochs, L. G. (1989). Toward the development of an elementary teacher's science teaching efficacy belief instrument. Presented at the 62nd annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Rushton, E. A., & Reiss, M. J. (2021). Middle and high school science teacher identity considered through the lens of the social identity approach: A systematic review of the literature. *Studies in Science Education, 57*(2), 141-203.

Rutherford, T., Liu, A. S., & Wagemaker, M. (2021). “I Chose Math Because…”: Cognitive interviews of a motivation measure. *Contemporary Educational Psychology, 66*, 101992.

Rutherford, T., Long, J. J., & Farkas, G. (2017). Teacher value for professional development, self-efficacy, and student outcomes within a digital mathematics intervention. *Contemporary Educational Psychology, 51*, 22-36.

Smolleck, L. D., Zembal-Saul, C., & Yoder, E. P. (2006). The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry. *Journal of Science Teacher Education*, *17*, 137-163.

Stains, M., & Vickrey, T. (2017). Fidelity of Implementation: An Overlooked Yet Critical Construct to Establish Effectiveness of Evidence-Based Instructional Practices. CBE life sciences education, 16(1), rm1. https://doi.org/10.1187/cbe.16-03-0113

Steyn, G.M. (2017). Fostering Teachers’ Professional Development Through Collaboration in Professional Learning Communities. In: Amzat, I., Valdez, N. (eds) Teacher Empowerment Toward Professional Development and Practices. Springer, Singapore. https://doi.org/10.1007/978-981-10-4151-8_16

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.

Unfried, A., Rachmatullah, A., Alexander, A., & Wiebe, E. (2022). An alternative to STEBI-A: Validation of the T-STEM science scale. *International Journal of STEM Education, 9*(1), 24.

Zhai, Y., Tripp, J., & Liu, X. (2024). Science teacher identity research: a scoping literature review. *International Journal of STEM Education, 11*(1), 20.

Acknowledgements

The authors would like to express their gratitude to the Practical Education Network (PEN) team for their thoughtful engagement and collaboration throughout this project. Their openness to research and willingness to share insights were invaluable in shaping the deliverables presented in this report. Special thanks go to Dr. Heather Beem, CEO and Founder, and Mawuena Asem Hanson, Research, Monitoring, Evaluation, and Learning Officer, for their expertise and contributions, which provided essential context and direction.

The Fellows Team also wishes to extend appreciation to the MIT Solve Team, especially Samantha Alper and Ellie Moriearty, for their guidance and support, which played a critical role in the success of this project.