The Journal of Mathematics and Science:

COLLABORATIVE EXPLORATIONS

Volume 19
Summer 2024

REGULAR JOURNAL FEATURES

Virginia Mathematics and Science Coalition
JOURNAL OF MATHEMATICS AND SCIENCE: COLLABORATIVE EXPLORATIONS

EDITOR
Aimee J. Ellington
Virginia Commonwealth University

FOUNDING EDITOR
P. N. Raychowdhury
Virginia Commonwealth University

EDITORIAL BOARD

Courtney Baker
George Mason University

Jessica Cohen
Western Washington University

Darryl Corey
Radford University

Julia Cothron
Virginia Mathematics and Science Coalition

Tom Dick
Oregon State University

Anthony Dove
Radford University

Elizabeth Edmondson
Virginia Commonwealth University

Diana Erchick
Ohio State University at Newark

William Haver
Virginia Commonwealth University

Shiv Khanna
Virginia Commonwealth University

Jim Lewis
University of Nebraska – Lincoln

Visala (Rani) Satyam
Virginia Commonwealth University

Toni Sorrell
Longwood University

Elizabeth Teles
Teles Consulting

Lindsay Wheeler
University of Virginia
The *Journal of Mathematics and Science: Collaborative Explorations* is the official journal of the Virginia Mathematics and Science Coalition ([https://www.vamsc.org/](https://www.vamsc.org/)), a non-profit organization comprised of education, scientific, corporate, and public policy leaders committed to the sustained elevation of mathematics and science education to ensure that all Virginia’s students and citizens have the foundation required for life-long success in their daily lives, careers, and society.

**AIMS AND SCOPE**

The *Journal of Mathematics and Science: Collaborative Explorations* is a forum which focuses on the exchange of ideas, primarily among higher education faculty from mathematics, science, and education, while also incorporating the perspectives of elementary and secondary school teachers. Articles are solicited that address the preparation of prospective teachers of mathematics and science in grades K–12, the preparation of mathematics and science teacher leaders for grades K–12, and innovative programs for undergraduate STEM majors.

Articles are solicited in the following areas:

- all aspects of undergraduate STEM education with particular interest in activities that will provide new insights in mathematics and science education
- all aspects of the preparation of mathematics and science teacher leaders and their work in K–12 schools and school districts
- reports on new curricular development and adaptations of “best practices” in new situations with particular interest in interdisciplinary approaches
- explorations of innovative and effective student teaching/practicum approaches
- research on student learning
- reports on STEM education projects that include evaluation
- reports on systemic curricular development activities in mathematics and science

To submit an article for consideration, go to [https://scholarscompass.vcu.edu/jmsce_vamsc](https://scholarscompass.vcu.edu/jmsce_vamsc)
# Table of Contents

## Regular Journal Features

**Reconciling Calculus Students’ Understanding of Average Across Multiple Contexts**
- F. Yu

**Seven Properties of Highly Effective Problems**
- T. Ales, K. Peterson and C. Roussos

**Two Decades of Supporting Excellence in STEM through Programs that Work: A History of High-Quality STEM Programming in the Commonwealth of Virginia**
- W. Haver and D. Neely-Fisher

**Introducing Flexible Assessment into a Computer Networks Course: A Case Study**
- J. Meehean

**Collaborative Leadership for Research Investigating STEM Teacher Preparation Across Many Institutions**
- D. Franz, K. Hubbard and D. Brenner
RECONCILING CALCULUS STUDENTS’ UNDERSTANDING OF AVERAGE ACROSS MULTIPLE CONTEXTS

Franklin Yu
Virginia Commonwealth University
yuf3@vcu.edu

https://doi.org/10.25891/hpmx-jx94

ABSTRACT
The idea of average is utilized in a variety of scenarios, yet the literature has indicated that students have multiple disconnected understandings for the concept of average. In this study, I provide an account of two students who reconciled their meanings for average by considering an average as a replacement with a constant value. This report discusses an intervention that teachers can leverage to help their students make their meanings for average coherent and conceptually based.

KEYWORDS
average, teaching intervention, quantitative reasoning
According to researchers, the concept of “average” is introduced early on, yet has often been cited as being difficult for students (Mokros & Russell, 1995; Watson & Moritz, 2000). By the time students get to Calculus, they have likely been exposed to the idea of “average” in statistical situations (e.g., finding the average test score), as an average rate of change, and even colloquially (e.g., the “average” person) (Bakker, 2003). However, studies have reported that students have conflicting meanings about average (Byerley et al., 2000; Watson & Mortiz, 2000) that are generally focused on a specific procedure. These studies indicate that students may not have constructed a meaning for average that conceptually connects these situations as coherent and related.

In the context of Calculus, it is important that students have a productive meaning for “average” when interpreting an average rate of change in a small input interval as an instantaneous rate of change (see the Conceptual Analysis of Average section below). However, researchers have indicated that most Calculus students’ meanings for average rate of change are often at odds with the idea of derivative as an imagined constant rate of change over an interval (Ärlebäck et al., 2013; Byerley & Thompson, 2017; Dorko & Weber, 2013; Frank & Thompson, 2021). These researchers noted that students might have issues with the concept of average rate of change since the word “average” has “lexical ambiguity because of its use in statistics and everyday language” (Dorko & Weber, 2013, p. 386). For example, Byerley and Thompson (2017) observed a student explaining the use of division in calculating a slope as “the division can be used because we talk about the slope being the average rate of change and ‘average’ is the total of the observations divided by the number of observations” (p. 188). Furthermore, some students confuse the idea of average with measures of central tendency (e.g., the median) (Zawojewski et al., 1997). The findings of these studies indicate that students need the opportunity to reconcile their understanding of the concept of average across these situations in order to develop a more coherent sense of mathematics. This study provides an example of how to support Calculus students in making such reconciliations.

**Theoretical Background**

To frame this study, I explicate the theoretical lens of quantitative reasoning (Smith & Thompson, 2007). According to Smith and Thompson, quantitative reasoning examines the thinking involved in conceptualizing a situation and its quantities. A quantity is a conceived attribute of a perceived object that someone envisions having a measurement. In this study, I leverage quantitative reasoning to focus on how an individual considers what the value of an “average” measures about a particular situation.

**Conceptual Analysis of Average**

Mathematics students’ first exposure to the idea of average is usually situated in discrete problems such as finding the average test score or the average height of the students in their classroom. In these contexts, students associate the idea of average with a specific procedure that entails adding up all the values and dividing the total by the number of items. However, researchers have reported that students have developed incomplete ideas of average as the middle of the data that can represent the set of data, as a number that provides a reasonable sense of the values of the data, and the number that is placed in the data such that the values higher and lower balance out (Mokros & Russell, 1995). These findings evidence that students often do not
form a quantitative meaning for what an average value represents, nor do they have the opportunities to understand why the procedure of adding and dividing produces the average. One productive understanding of the average value in these situations would be the idea of replacement with a constant value where one asks, “What constant value would I need such that we obtain the same total as the original data set?” This explains the procedure of adding up (finding the total) and dividing by the number of items (redistributing the total into equally sized portions).

In the context of Pre-Calculus and Calculus, we can consider an Average Rate of Change as a hypothetical constant rate of change over a function’s input interval that achieves the same change in the output quantity, as achieved by the function, over the input interval, from \(x_1\) to \(x_2\), on which the average rate of change is determined (i.e., \(\frac{f(x_2)-f(x_1)}{x_2-x_1} = m\), where \(m\) is the value of the imagined constant rate of change). We can continue to utilize the idea of average here as a replacement with a constant value by thinking of an average rate of change as “the constant rate of change that would be needed for the output value to change the same amount over the same input interval.” This way of thinking also explains the reason for a different procedure since we are attending to changes in two quantities’ values (rather than a sum) in this context.

**Methodology**

This study was part of a larger study that was conducted by engaging students in individual teaching experiments (Steffe & Thompson, 2000). The students selected were enrolled in a Calculus 1 for Engineers course before learning about secant lines and instantaneous rate of change. The teaching experiment involved six sessions (1 pre-interview and 5 teaching sessions) that focused on characterizing and advancing students’ ways of thinking about rate of change. This paper reports on two of the students in a single teaching session where the teacher worked with the students to develop a productive understanding of average rate of change in the context of Calculus.

**Background**

The teaching experiment was designed to support students in developing a productive meaning for instantaneous rate of change (e.g., the derivative) based in quantitative and covariational reasoning (Carlson et al., 2002). In the first session (the one prior to one discussed in this paper), students worked through a Desmos applet to develop a meaning for constant rate of change that involved the constant ratio between changes in two quantities values (Yu, 2023). As students progressed through the teaching experiment, they would continue to build on their meaning for constant rate of change into contexts of average rate of change and then an average rate of change over smaller and smaller intervals.

Table 1 provides an overview of the second teaching session, where the students had to calculate an average speed and explain what having an average speed represented in the context. In particular, the latter portion of this task anticipates that the students likely have not yet developed a productive quantitative understanding of average rate of change and provides a small intervention to support students in constructing one.
### Table 1
**Teaching Session 2 - A Teaching Intervention for Average Rate of Change**

<table>
<thead>
<tr>
<th>Average Rate of Change – The Runner Task</th>
<th>Task Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://www.desmos.com/calculator/bpdeilsrshb">https://www.desmos.com/calculator/bpdeilsrshb</a></td>
<td>This portion of the task is to elicit students’ understandings about average rate of change and how to calculate it. During this portion, the instructor asks students what they are writing and what they believe is being represented. This is to have students attend to variations in quantities and engage in using division to represent a ratio.</td>
</tr>
<tr>
<td>Jonah is running on a racetrack that is 100 meters long and his distance from the starting line is given by the function $s$, where $s(t)$ represents his distance in meters from his starting spot after running for $t$ seconds. Jonah finishes running the 100 meters in 32 seconds.</td>
<td></td>
</tr>
<tr>
<td>a) Does Jonah run at a constant rate of change (with respect to time)? How do you know? (What would you do to verify your answer.)</td>
<td></td>
</tr>
<tr>
<td>b) What was his average speed? How would you represent his average speed in the first 23 seconds of his run? Between 4 seconds and 8.9 seconds?</td>
<td></td>
</tr>
<tr>
<td>Suppose we wanted to run the same race as Jonah. We want to travel the same total distance and use the same amount of time as Jonah did. However, we want to run at a constant speed. What constant speed would that have to be? [Students can put their answer into the Desmos file and run the animation to check.]</td>
<td>This portion of the task is to have students continue to engage in reasoning about constant rate of change.</td>
</tr>
<tr>
<td>The calculation you did for finding the constant speed is the same as for the average speed of the 1st runner over the entire 32 seconds. Why is that? What does average speed mean?</td>
<td>This portion of the task promotes that an average rate of change involves a constant rate of change (not necessarily the entire definition of an average rate of change, instead this task provides an opportunity for students to make a connection that an average speed entails a constant speed).</td>
</tr>
<tr>
<td>(Optional Task) Suppose out of 5 quizzes (graded out of 10 points) you earned a 7, 9, 10, 4, 9. What is your average test score?</td>
<td>Students may not have strong quantitative meanings for average outside of “add up everything and divide.” I provide additional examples to help students realize that “average” is about a replacement with a constant.</td>
</tr>
<tr>
<td>If you earned the same total score as the above and scored the same score on each quiz, what score would that have to be?</td>
<td>I designed this optional task to perturb students who rely on calculating an average as “add up everything and divide” by having them reconcile the meaning of “average” in one context with another. I intend for students to reflect on their meaning of average in Task 3 and consider that an average involves a replacement of values. (For the previous task, one meaning may be to replace all of the 1st runner’s speeds with one speed that would have him travel the same distance in the same amount of time).</td>
</tr>
<tr>
<td>How is “average” here similar to “average” in average speed?</td>
<td></td>
</tr>
</tbody>
</table>

**Summary of Session 2:** The intent of this task is to explore students’ meanings for average rate of change and potentially perturb their understandings of average as “add up and divide”. The goal is to support students in interpreting “average” to mean a replacement of values with a constant value. In the context of Average Rate of Change, a student is determining variations in two quantities and then finds a ratio between them. Then the student uses the value of the ratio to imagine how the values of the quantities will vary if they were to vary at that constant rate. I do not intend for this task to be a learning trajectory on average rate of change. The entire teaching experiment was designed to be a learning trajectory for instantaneous rate of change that would be feasible to conduct within the time constraints of an undergraduate Calculus 1 course.
Results

During the Runner Task, one of the students named Scott stated that $\frac{s(32) - s(0)}{32 - 0}$ was Jonah’s average speed since it was a change in distance over a change in time. However, he also verbalized that an average is “every different speed that he’s moving at, add them all together and divide and call that the average.” Further, he stated, “like two different points from where he is running and we find the average slope,” and he drew a line between two points on a graph (see Figure 1). While Scott’s explanations contradicted each other, his interpretation represented a common understanding of “average” across different contexts. Scott expressed three inconsistent meanings for average rate of change (i) a procedural meaning as a change in output values divided by a change in input values, (ii) a process meaning as “add up and divide,” and (iii) an “average slope” obtained by determining the slope of a line between two points on a function’s graph. As Scott discussed his understanding of average in different contexts, it did not seem like Scott recognized the inconsistencies in these conceptions. Scott had disconnected understandings of average rate of change which parallels how some students develop disconnected understandings of derivative (Zandieh, 2000)\(^1\). Hypothetically, it stands to reason that if these inconsistencies were not addressed, Scott might have compartmentalized different understandings for derivative: procedurally as the limit of the difference quotient, a process of calculating limits or sliding a secant line, and graphically as the slope of a tangent line.

Figure 1
Scott’s Drawing of a Secant Line with an Average Slope

To perturb Scott’s way of thinking, the interviewer introduced a question about an average quiz score and asked what the average quiz score would describe in that situation (see Figure 2). Scott replied, “I have no idea what to say because average is one of those words that I

Figure 2
The Quiz Problem

\(^1\) Zandieh (2000) indicated that there are several (often disconnected) representations of derivatives that students tend to recall: graphically as the slope of the tangent line, verbally as instantaneous rate of change, physically as speed, and symbolically via the limit definition of derivative, yet some students are often unaware of how and why these ideas are related to each other.
use in a lot of other situations…like here it might be like the middle score”. The interviewer then asked Scott to consider two of his meanings for average simultaneously by writing out the calculations for an average for the Runner Task and the quiz problem (see Figure 3). Next, the interviewer provided a hypothetical second student who scored 8 on all five quizzes and highlighted that both students had the same total quiz score (see Figure 4). Scott remarked, “it’s kind of clicking for the other one [The Runner Task], I guess like they both run the same total they just run different speeds throughout that total distance… like they both run the same total distance for the same total time with different speeds.”

**Figure 3**
*Representation of “Average” in the Runner Task and Quiz Problem*

![Figure 3](image1.png)

**Figure 4**
*Average Quiz Score as the Constant Score Needed for all 5 Quizzes to have the Same Total Quiz Score for all 5 Quizzes*

![Figure 4](image2.png)

When presented with the quiz score problem, Scott verbalized his awareness that he had different meanings for average, depending on the context. Having Scott consider the juxtaposition between the average speed of the runner and the average quiz score supported Scott in realizing the inconsistencies across his meanings for average. Prompting Scott to compare the runner and quiz situations caused him to consider an average as involving the same net changes over the same input interval in each situation. For example, in Figure 3, Scott wrote $\frac{s(32) - s(0)}{32-0}$ to
represent Jonah’s average speed and \( \frac{5+6+9+7+10}{5} \) to represent the average quiz score, and seeing both as representing an “average” deterred him from explaining an average as a particular procedure. Instead, Scott noticed that “they both run the same total for the same total time in the runner task.”

In the next session, Scott was asked what he recalled about the Runner Task. Scott mentioned the two runners, one running at a constant speed and the other at a varying speed, but both finished running 100 meters in the same time. He then verbalized that the runner who had an average speed of \( \frac{100}{32} \) meters per second meant that “if every value of speed were to be the same, they would still travel the 100 meters in 32 seconds.” Scott’s recollection indicated a shift in his understanding of average speed as entailing the same total distance and the same amount of time as someone who would travel at a constant speed of \( \frac{100}{32} \) meters per second for 32 seconds.

In the Runner Task, another student, Hans, interpreted an average speed in a similar manner as Scott. Hans also recognized that \( \frac{s(32) - s(0)}{32 - 0} \) was “Jonah’s average constant speed” since it was a change in distance divided by a change in time (see the left side of Figure 5). When asked to explain what he meant by average, he struggled to verbalize his meaning. He eventually drew a graph of the situation (Jonah’s distance traveled with respect to time elapsed) and a dashed blue secant line and then indicated that the whole line was the average (see the right side Figure 5).

Figure 5
Hans’ Interpretation of \( \frac{s(32) - s(0)}{32 - 0} \) and Drawing of a Secant Line as the Average

Unlike Scott, Hans never mentioned that his interpretation of average speed included a process of adding up and dividing. However, Hans’ interpretation of the secant line as representing the average speed indicated his lack of associating quantities with features of a graph. (It should be noted that Hans also did this in previous teaching sessions.) The interviewer then recreated Hans’ drawing and asked him what a point on his secant line represented in the runner context (see the left side of Figure 6). After determining that a point on the secant line would be an ordered pair of time elapsed and distance traveled, Hans said that the average speed would be the constant speed over the time period between 0 and 32 seconds. He then drew the associated total distance and time on the graph to match what he said (see pink lines on right side of Figure 6). After drawing, he paused and noted that the average speed of \( \frac{100}{32} \) meters per second
would mean that “Jonah [the first runner] will end at the same distance and time as Ishtesa [the constant speed runner]”; however, he said that he still did not know how to explain what \( \frac{100}{32} \) quantified about Jonah.

**Figure 6**
Hans’ Interpretation of Secant Line and Observation of the Total Distance and Total Time in the Runner Task

To support Hans in developing a meaning for the value of an average speed, the interviewer employed the same average quiz score situation used in Scott’s teaching session. After working through the example and considering both the runner and quiz score scenarios (see Figure 7), Hans indicated that “they [the situations] were similar because like the totals were the same, like on the left [the runner situation], Jonah has the same ratio of distance and time, and on the right [quiz score situation] the students have the same ratio of scores and tests.” Hans was then prompted to explain what Jonah’s average speed of \( \frac{100}{32} \) meters per second meant, and he updated his description to “Jonah having an average speed of \( \frac{100}{32} \) means that if his speed fluctuates, he will end at the same time and distance as if he were going at a constant rate.”

**Figure 7**
Hans Comparing the Runner Task and the Quiz Problem
After conducting the retrospective analysis of this session, it is likely that having Hans attend to the quantities that composed an average (the total distance and total time in the runner task, and the total quiz score and total number of quizzes in the quiz problem) helped him articulate a productive understanding for average rate of change. As the session progressed, Hans’ thinking about an average shifted from a secant line connecting two points on a graph (see right side of Figure 5) to interpreting Jonah’s average speed as involving the same total distance and total time (see Figure 7), and finally, that the value of an average speed involved a hypothetical constant rate of change. What supported Hans’ actions was the prompting to associate numbers and features of a graph with quantities. For instance, Hans initially thought the secant line represented the average speed. After being asked what a point on that secant line represented (see left side of Figure 6), he realized the average speed involved the total distance and time traveled (see right side of Figure 6). Later, Hans noticed the similarities across the runner and quiz score problems by taking note of the same ratios in each context. He then connected how ‘average’ referred to this similarity by updating his description of Jonah’s average speed to entail the constant speed that resulted in the runner traveling the same total distance over the same period of time. This description provided evidence of a shift in Hans’ understanding of average speed.

Discussion

Reminiscent of the literature, both Scott and Hans initially demonstrated an incomplete understanding of average. With both students, it was productive to support them in attending to what the numerical value of an average represented and how it related to particular quantities within a situation. Additionally, having them reconcile their disconnected meanings for average (particularly for Scott) by evoking the term “average” in different situations supported them in developing a meaning for average based in quantitative reasoning. I argue that these two pieces of (1) having students attend to conceptual understandings and (2) seeing two different examples of “average” being used simultaneously can be powerful in shifting thinking and better preparing them for developing a meaning for instantaneous rate of change (i.e., an average rate of change in a small interval). In this study, both students benefited when comparing the average quiz score situation with an average rate of change in the runner problem due to having to reflect on why “average” could be appropriately applied in both contexts. This type of intervention could be used by other instructors when revisiting average rate of change (slopes of secant lines) before discussing instantaneous rate of change (slopes of tangent lines) to support students in quantitative reasoning. In particular, I assert that students need the opportunities to build such conceptual understandings by reflecting on their current meanings and how they may not be entirely coherent.

While the findings of this study can provide insight into supporting students’ development of average rate of change, the findings can only be said to be true for the two students presented in this paper. However, despite this, this idea of reconciling disconnected meanings through reflection is theoretically supported by Radical Constructivism (Thompson, 2000) and the works of Piaget (1971). One teaching implication is that Calculus instructors should deliberately spend time reconciling students understanding of average rate of change prior to formally introducing the derivative via the limit definition. If instructors aid students in developing quantitative meanings for rate of change, students will be less likely to form disconnected understandings of the derivative concept (Zandieh, 2000), and can instead
understand that Calculus is the mathematics of modeling how quantities covary. Future studies can further investigate the implementation of such an intervention within a Pre-Calculus or Calculus course.

References


SEVEN PROPERTIES OF HIGHLY EFFECTIVE PROBLEMS

Thomas Ales  
*University of Lynchburg*  
ales_tm@lynchburg.edu

Kevin Peterson  
*University of Lynchburg*  
peterson@lynchburg.edu

Constantine Roussos  
*University of Lynchburg*  
roussos@lynchburg.edu

ABSTRACT
In an effort to provide more critical thinking opportunities in their courses, instructors are embracing the power of problem- and project-based learning (PBL). In this paper we address the importance of problem quality when utilizing PBL. We list seven important properties that a high-quality problem should have. We conclude with an example of a problem that possesses all seven properties.

KEYWORDS
problem-based learning, project-based learning, experiential learning, mathematics
A goal of educators is to give their students an opportunity to think critically. According to Google's English dictionary (provided by Oxford Languages) the definition of critical thinking is “the objective analysis and evaluation of an issue in order to form a judgment.” The example sentence provided with the definition is surprising: "Professors often find it difficult to encourage critical thinking amongst their students." Thus, the motivation for this paper is understanding how educators can create an environment where students feel comfortable thinking critically.

One popular method is using Problem-based Learning (PBL). A search for this phrase in the digital library JSTOR returns over 700,000 articles and book chapters. These articles describe studies of many aspects of PBL but only one review paper by Yew and Goh (2016) mentioned anything about the effect of problem quality on student learning. We have found that an important part of “teaching” critical thinking is identifying problems that will hold a student’s attention and inspire them to persevere. It is an attractive idea to have a student work on a problem of their own devising so that they are “invested” in the experience. Unfortunately, student designed problems are often poorly defined and very difficult to solve. And personal experience tells us that a lot of effort spent on an unyielding problem does not encourage continued engagement and interest. When teaching critical thinking, we need to avoid situations where students feel there is no possibility of success. Hence, it is crucial to assign problems so that students have a “curated” critical thinking experience that strengthens both their mathematical knowledge and their intellectual confidence. Additionally, these curated experiences are less stress on the instructor because the educator knows what the result will be and is better able to provide the appropriate level of support required for each individual student.

The Setting

The majority of the problems and examples mentioned in this paper have been used in undergraduate courses taught at the University of Lynchburg. Over the past 23 years, we have developed, taught, and enhanced several project-based courses. Below we list a few of these courses with a brief description.

- Problem Solving in Mathematics (MATH 105) is a general education course that introduces students to the true problem-solving nature of mathematics. The focus is on using quantitative reasoning and intuitive logical thought techniques to solve problems rather than formal rigid processes.
- The Mathematics of Computer Science (MATH 231) introduces the theoretical and mathematical foundations of computer science through a combination of traditional lectures, problem sets, and six individual projects.
- Experimental Mathematics (MATH 350) is a project-based course that introduces students to the fine art of problem solving. The focus is on using computers, models, and examples to investigate problems to uncover possible solutions.
- Discrete Mathematics (MATH 330) covers topics such as counting, graph theory and cellular automata. It is a blend of traditional lectures, problem sets, and individual projects.
- Senior Research (MATH 451 and STAT 451) are the capstone courses for students majoring in mathematics or statistics. Students work on at least one major project throughout the semester. These projects allow students to dig deeper into key topics from earlier in the curriculum. Each student must write a research paper and present their work to the class.
• Statistical Methods (STAT 400) uses projects to explore topics such as estimation, inference, comparative analysis, analysis of variance, multiple comparisons, regression, analysis of covariance, and measures of fit regarding multiple regression models.

In addition, two of the authors, Ales and Peterson, have taught experiential-learning, project-based courses at Virginia’s Summer Residential Governor's School for Mathematics, Science, and Technology, which provide gifted high school seniors with immersive experiences in what it really means to be a mathematician.

All of these courses have a strong PBL component that involves weekly individual or group meetings with the students. All student anecdotal comments and feelings mentioned below come from conversations that took place during these meetings. Our extensive experience has motivated us to develop the following seven properties of problems that inspire critical thinking.

Property 1 – Easy to Understand

There is a time and a place for complicated problem statements that force students to parse through the combination of language, variables, and notations to determine what the question is asking. Challenges such as these are excellent for the seasoned problem solver, but will rarely entice the average student to dive into a problem. This problem from the 1962 International Mathematical Olympiad (IMO Board, n.d.) is an excellent example of what to avoid when selecting a problem:

Consider the cube $ABCDA'B'C'D'$ where $ABCD$ and $A'B'C'D'$ are the upper and lower bases, respectively, and edges $AA', BB', CC', DD'$ are parallel. The point $X$ moves at constant speed along the perimeter of the square $ABCD$ in the direction $ABCDA$, and the point $Y$ moves at the same rate along the perimeter of the square $B'C'CB$ in the direction $B'C'CBB'$. Points $X$ and $Y$ begin their motion at the same instant from the starting positions $A$ and $B'$, respectively. Determine and draw the locus of the midpoints of the segments $XY$.

Of course, the simplicity of the problem statement does not imply that the problem is simple to solve but a well-worded, easy to understand problem is less likely to intimidate students before they attempt to solve it.

Property 2 – The Result is not Given

The student should have to make a conjecture. There is plenty of room in the curriculum for problems that say, “prove the following result”, but we have found that there is more interest if the result is not given. Instead, create a problem that asks students to explore a mathematical situation and make a conjecture on their own. For instance, a problem like:

Find a formula for $1 + 3 + 5 + \cdots + 2n - 1$ that depends on $n$ and prove that your conjecture is correct.

This problem encourages experimentation and has the potential to reveal a surprising solution, where a problem like:
Prove: \(2 + 4 + 6 + \cdots + 2n = n(n + 1)\)

has already revealed its surprise and only serves as an exercise in inductive techniques. Everyone loves a treasure hunt and much mathematics is learned by constructing examples and searching for patterns. This is the heart of mathematics!

**Property 3 – Easy to Do Examples**

The problem should be easy to model, by hand or by writing code, and the results of this modeling should be used to make or check the conjecture. An important part of problem solving is understanding the problem (Polya, 1957) and generating multiple examples is a great way to get a better feel for what is happening in a given problem.

When students are looking for a conjecture for the sum of the first \(n\) odds, for example, they almost immediately start writing down examples and searching for a pattern. Most students create something similar to the following table:

<table>
<thead>
<tr>
<th>(n)</th>
<th>Sum (= 1 + 3 + 5 + \cdots + 2n - 1)</th>
<th>Result (= n^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1 + 3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1 + 3 + 5</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>1 + 3 + 5 + 7</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>1 + 3 + 5 + 7 + 9</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>1 + 3 + 5 + 7 + 9 + 11</td>
<td>36</td>
</tr>
</tbody>
</table>

After these examples, students are ready to make the conjecture that \(1 + 3 + 5 + \cdots + 2n - 1 = n^2\). In fact, through this modeling, they also develop other insights into the sequence of squares. Generating examples gives students a chance to think algorithmically, gives them a way to check their result, and gives them an opportunity to move from specific examples to the general result.

**Property 4 – Know the Solution**

The problem should be tractable and the instructor should know the solution. If we do not know the solution, it is difficult to properly advise students or give good hints. Students can get frustrated without the appropriate support. An example of a problem that is easy to understand and students find interesting is the Collatz conjecture or the \(3x + 1\) problem which is stated as follows:

Define a sequence \(a_k\) in the following way: \(a_1\) is any positive integer and

\[
 a_{n+1} = \begin{cases} 
 \frac{a_n}{2} & \text{if } a_n \text{ is even} \\
 3a_n + 1 & \text{if } a_n \text{ is odd}
\end{cases}
\]

Collatz states that this sequence eventually reaches 1 regardless of the starting integer (Lagarias, 2010).
When asked about the Collatz conjecture, Paul Erdös stated “[m]athematics is not yet ready for such problems” (Guy, 2004, p. 330). Jeffrey Lagarias, a University of Michigan professor who has written extensively about this problem, stated that the Collatz conjecture “is an extraordinarily difficult problem, completely out of reach of present-day mathematics” (Lagarias, 2010, p. 4). Clearly, this makes it quite difficult to properly aid students in pursuit of this problem. We think Randall Munroe’s (n.d.) website XKCD.com (see Figure 1) best describes what happens in these cases.

**Figure 1**

*XKCD Comic about Collatz Conjecture*

When instructors know the solution, they can more easily provide the support needed by each individual level of student. This support is the basis of the curated critical thinking experience.

**Property 5 – From the World**

The problem comes from a real-world situation or a game. It is motivating to have a tangible reason to study a problem, even if the implied connection is a bit farfetched. We have used the following three versions of the same problem in the Mathematics of Computer Science (MATH 231). Here is the original version without examples and extra explanations:

We begin with an $n \times n$ grid with each cell containing either a 1 or a 0. We will call this the start configuration. Our simple rule for this automaton is that at each time step any cell containing a 0 that has at least two neighbors that contain a 1, becomes a 1. Here a neighbor is directly to the left, right, above or below a given cell, not diagonal. What is the least number of 1’s required in a start configuration to guarantee that the grid will eventually fill with 1’s?
Students found this problem somewhat interesting. During our first attempt at using this problem as part of a class activity, students needed to come up with three related problems. We found that this problem wasn’t very inspiring for them.

The next year, we changed the problem to an $n \times n$ grid of trees where some of the trees were on fire. The same rule as described above represented the “spreading” of the fire. There was certainly more interest in this version, and, as a result, the students proposed more interesting “new” problems.

Most recently, we have changed the problem from an $n \times n$ grid of trees to an $n \times n$ classroom of students. The goal in this scenario was to create a simple model of the spread of COVID-19. The problem situation starts with some infected students in the classroom and followed the same rules as outlined above. The students were very interested in this model. The problems proposed by students were more interesting and often included the obvious addition of probabilities to our rules for “infection”. That is, they naturally changed the problem from being automatically infected with two infected neighbors to a probabilistic rule that depended on the number of infected neighbors. There is true value to changing the statement of a problem to make it more applicable to how the students are experiencing the world around them while maintaining the problem’s mathematical integrity.

**Property 6 – Easy to Generalize**

Interesting problems naturally generate more problems. We are not interested in a “one and done” problem, but are instead looking for problems that have many (fairly) obvious, interesting modifications. One of our favorite problems is from a book called *Dueling Idiots and Other Probability Puzzlers* (Nahin, 2012). We modified the dueling scenario slightly as follows and call our version of the game Bagsy:

Consider a simple two-player game where the numbers 1 – 6 are written on six ping-pong balls with one number on each ball. The numbered balls are placed in a bag. Player 1 reaches into the bag (without looking) and removes a ball. If that ball is numbered “1”, then Player 1 wins. If it is any other ball, Player 1 puts the ball back in the bag, shakes it and hands it to Player 2 who repeats the process. The first player to draw the ball numbered “1” wins.

The question is: What is the probability that Player 1 wins? Clearly, this problem has a plethora of interesting modifications that will change the probability that Player 1 wins. We mention the two most obvious: First, simply change the number of balls in the bag to $n$, and second, change the number of players to $k$. These generalizations give students the opportunity to think like a mathematician. The mathematics a problem generates can continue to engage the student long after the original problem is solved.

**Property 7 – The Problem Contains a Surprising Result**

Typically, the surprise when solving a mathematical problem in a class is that the proof requires some technique or result from a previous lesson or a previous course. We again use Bagsy (described above) as an example. In the original version of the game, it should be clear that it could take an arbitrarily long time to finish the game. In fact, the expression for the
probability that Player 1 wins is an infinite series. The nice surprise is that this series is a
convergent infinite geometric series, and we can use well-known results from Calculus II to
arrive at the answer. These results, often from previous courses, highlight the connectedness of
mathematics.

Over years of teaching project-based courses and developing experiential learning
modules we have found that problems with these seven properties do an excellent job of keeping
a student’s attention while inspiring voluntary critical thought.

Problem Example

We now give an example of an interesting problem (with a solution) that has all the
properties mentioned above. This problem developed from a daily task of one of the authors.
When walking their dog, they typically take a handful of pistachio nuts to snack on during the
outing. As anyone who has eaten pistachios knows, there are two kinds of pistachios: Type 1 are
those that have a split in their shell that makes them easy to open and Type 2 are those that have
no split and require reconstructive dental work after attempting to open. During the walk, the
author would reach in their pocket and pull out a random pistachio. If it is a Type 1 pistachio,
they open it, eat it and toss the compostable shells. If it is a Type 2 pistachio, they return it to
the same pocket, mix up the nuts and choose a new one. The following question arises: On average,
how many times is a Type 2 pistachio chosen before all the Type 1 pistachios are consumed?

To make the problem simpler to talk about, we turn it into a game of chance. Imagine the
following game:

Fill a bag with ten black marbles and one red marble, shake the bag and choose
one marble without looking. If that marble is black, remove it from the bag and
choose again. If the chosen marble is red, collect one dollar, return the red marble
to the bag and choose again. The goal is to accrue as much cash as possible before
all the black marbles have been chosen.

The question is: On average, how much money would you expect to win per game?

Let \( E(n) \) represent the expected winnings when the bag contains \( n - 1 \) black marbles
and one red marble. The formula for this quantity is quite simple (the term in the sum for
selecting zero red marbles is omitted):

\[
E(n) = 1 \cdot P(\text{picking red exactly once}) + 2 \cdot P(\text{picking red exactly twice}) + \cdots \tag{1}
\]

Clearly, each game can be arbitrarily long. A very lucky (and unlikely) player may continue to
choose only the red marble indefinitely, meaning the expression for \( E(n) \) is an infinite series.

We wrote code (using MATLAB) to model this game. We were curious if there were any
obvious patterns in the experimental results for different values of \( n \), and wanted to have data to
check conjectures. The code generated results for \( n = 2 \) to 16 marbles (one of which is red). The
model “plays” the game 10,000 times for each \( n \) and calculates the average winnings. The table
of results is provided below.
### Table 2

<table>
<thead>
<tr>
<th>$n$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(n)$</td>
<td>0.99</td>
<td>1.5</td>
<td>1.8</td>
<td>2.1</td>
<td>2.3</td>
<td>2.5</td>
<td>2.6</td>
<td>2.8</td>
<td>2.9</td>
<td>3</td>
<td>3.1</td>
<td>3.2</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Initially, the only noticeable pattern is that the expected value grows very slowly.

Next, we attempted to calculate cases of $E(n)$ directly, but quickly recognized the folly in such an approach. Nevertheless, there is value in calculating $E(2)$ directly. Notice that when starting with two marbles (one black and one red) the game is over as soon as you choose the black marble. Furthermore, the probability of choosing the red marble exactly $k$ times before choosing the black marble is $\left(\frac{1}{2}\right)^k \frac{1}{2} = \left(\frac{1}{2}\right)^{k+1}$. This means $E(2)$ is

$$
E(2) = 1 \cdot \left(\frac{1}{2}\right)^2 + 2 \cdot \left(\frac{1}{2}\right)^3 + 3 \cdot \left(\frac{1}{2}\right)^4 + \cdots = \frac{1}{2} \sum_{i=1}^{\infty} \frac{i}{2^i} = 1
$$

This “scaled” geometric series is fairly well-known. Notice that the result matches our model quite nicely. It is left to the reader to calculate $E(3) = 1.5$. Attempts were made to compute $E(n)$ for higher values of $n$, but such calculations proved prohibitively difficult. A new approach was needed, but a conjecture was in sight.

Instead of calculating the probabilities directly, we tried to develop a simple recursive relationship by removing one marble. We split the problem into the two obvious parts; whether a red or black marble was chosen first. Clearly, with probability $\frac{1}{n}$ the red marble could be chosen and replaced or with probability $\frac{n-1}{n}$ a black marble could be chosen (and subsequently removed) leaving us with $n - 1$ marbles in the bag. This thinking yields the following expression as a conjecture for the expected value:

$$
E(n) = \frac{1}{n} (E(n) + 1) + \frac{n-1}{n} E(n-1)
$$

Solving for $E(n)$ in (3) we arrive at:

$$
E(n) = E(n-1) + \frac{1}{n-1}
$$

Using our result for $E(2)$ and the above recurrence relation, we arrive at the following surprising result:

$$
E(n) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}
$$

which is the partial sums of the harmonic series.

Before we prove that (3) is true, we will check to see how well our model matches the proposed solution found in (5). We display two graphs below that show the model and the proposed solution on the same graph for $2 - 100$ marbles. The first graph (Figure 2) was created using 1,000 trials in the model and the second (Figure 3) using 10,000 trials.
Now that we are convinced that this result is likely to be correct, we will prove that the recurrence relation found in (3) is true.

**Main Result**

The expected winnings, $E(n)$, when playing the game with $n - 1$ black marbles and 1 red marble satisfies the following recurrence relation $E(n) = \frac{1}{n} (E(n) + 1) + \frac{n-1}{n} E(n - 1)$. 

**Figure 2**

$E(n)$ Compared to Our Model with 1,000 Trials

**Figure 3**

$E(n)$ Compared to Our Model with 10,000 Trials
Proof:
We start with the definition for the expected winnings as given in (1).

\[ E(n) = \sum_{i=1}^{\infty} i \cdot P(i, n) \]

Here \( P(i, n) \) is the probability of drawing the red marble \( i \) times when there are \( n \) total marbles at the start. The individual probabilities \( P(i, n) \) can be broken down in the following way because drawing a red marble first and drawing a black marble first are mutually exclusive events:

\[
P(i, n) = P(i, n|\text{draw black first}) + P(i, n|\text{draw red first})
\]

\[
= \frac{n-1}{n} P(i, n-1) + \frac{1}{n} P(i-1, n)
\]

(6)

We now rewrite \( E(n) \) using (6), properties of summations, (1), \( i = i - 1 + 1 \), and the fact that for each \( n \) we have \( \sum_{i=0}^{\infty} P(i, n) = 1 \) to obtain the following:

\[
E(n) = \sum_{i=1}^{\infty} i \cdot \left( \frac{n-1}{n} P(i, n-1) + \frac{1}{n} P(i-1, n) \right)
\]

\[
= \frac{n-1}{n} \sum_{i=1}^{\infty} i \cdot P(i, n-1) + \frac{1}{n} \sum_{i=1}^{\infty} i \cdot P(i-1, n)
\]

\[
= \frac{n-1}{n} E(n-1) + \frac{1}{n} \left( \sum_{i=1}^{\infty} (i-1) \cdot P(i-1, n) + \sum_{i=1}^{\infty} P(i-1, n) \right)
\]

\[
= \frac{n-1}{n} E(n-1) + \frac{1}{n} \left( \sum_{j=0}^{\infty} j \cdot P(j, n) + \sum_{j=0}^{\infty} P(j, n) \right)
\]

\[
= \frac{n-1}{n} E(n-1) + \frac{1}{n} (E(n) + 1)
\]

which is the desired result.

Conclusion

This result is both surprising and satisfying. It is rare that a problem from the real world generates such an interesting solution. We see that this problem possesses all seven of the important properties required to entice critical thinking in students. It is certainly easy to understand, there is no conjecture provided, the problem is easily modeled either by writing code or by actually playing the game, the problem is accessible to both undergraduates and strong high school students with some understanding of probability and expected value, it is motivated from a real-world situation, has a surprising result (i.e., the solution is the partial sums of the harmonic series), and it is easily generalized.

Below you will find a few problems that are motivated by the original pistachio problem but stated in terms of colored marbles.

1. The most obvious generalization is to add more red marbles. This turns out to have a very nice solution as well. The solution is left to the reader.
2. We can add players and have them choose a marble in turns. The player that has the most money wins. Find the probability that Player 1 wins or find the expected value for each player.

3. Create a new game. Start with $n$ black marbles and $k$ red marbles and one player. The player chooses one marble at a time and removes it from the bag (regardless of color). The player wins if they are able to remove all the black marbles before removing all the red marbles. Find the probability that the player wins. This is very similar to the old Price is Right game “Strike Out”.

4. Play the same game as above but every time you choose a red marble, you remove it and add a black marble. The player wins if they are able to remove all the black marbles before removing all the red marbles. Find the probability that the player wins.

Hopefully, the reader can see that we could fill the page with interesting problems that are related to the original.

As you employ PBL in your classes, we hope you now have an appreciation for the importance of problem selection and how the right problems can help you help students acquire critical thinking skills. With this in mind, we end this article with a short list of interesting books that contain great examples of high-quality problems. We purposefully did not include most of Martin Gardner’s books, as most people are already familiar with those gems.

- Nahin, P. J. (2004). *When least is best: How mathematicians discovered many clever ways to make things as small (or as large) as possible*. Princeton University Press.

References


https://www.jstor.org/stable/j.ctt7ssjr


https://doi.org/10.1016/j.hpe.2016.01.004
ABSTRACT
The Virginia Mathematics and Science Coalition annually recognizes effective science, technology, engineering, and mathematics (STEM) programs. The leaders of these Programs That Work receive recognition and others gain ideas to incorporate into their STEM programs. Programs That Work was initiated in March 2000 as a part of a statewide conference designed to better understand effective strategies to increase the success of women, minorities and members of other groups who had been underrepresented in STEM. Programs That Work has since expanded to include effective STEM programming for all students and teachers at all levels by recognizing school systems, colleges and universities and other institutions, while maintaining a special emphasis on underrepresented groups. This paper summarizes the statewide conference and provides a description of the types of programs recognized over the last two decades. Lastly, the 2023 ceremony is described along with brief summaries of each of the seven recognized programs.

KEYWORDS
programs that work, success of women, recognizing effective programs, success of minorities
The Virginia Mathematics and Science Coalition (henceforth referred to as the Coalition) is committed to supporting high quality, effective STEM programs and student learning at all levels. One of the ways the Coalition provides this support is through Programs That Work. Annually, through this award program, the Coalition recognizes effective science, technology, engineering, and mathematics (STEM) programs for students and for teachers throughout the Commonwealth of Virginia. The leaders of these programs receive well deserved recognition and others gain ideas as they develop new STEM programs or as they refine their existing programs.

A Challenge to All: Raising the Participation and Success of Women and Minorities in Mathematics, Science, and Technology

Programs That Work was initiated in March 2000 as a part of a statewide conference attended by industry and organization leaders, educators, and other stakeholders to better understand effective strategies to increase the success of women, minorities and members of other groups who had been consistently underrepresented in STEM fields. During the past twenty-three years, Programs That Work has expanded to recognize STEM programming for all students from K-12 to higher education, as well as STEM professional development programs for teachers, while maintaining a special emphasis on underrepresented groups. The statewide conference was convened in Chester, Virginia, on March 23 – 24, 2000. Dominion Energy provided the space for the conference and was the major sponsor along with the Collins-Warner Foundation and the IBM Corporation. Coalition member LaRay Mason, Cindy Balderson from Dominion, and Coalition member and author Bill Haver organized the conference.

In the first plenary session, Coalition member, Donna Sterling of George Mason University presented research concerning the barriers that limit participation in STEM, including the almost universal neglect of girls during classroom question-and-answer sessions. The findings were that female students were called upon much less frequently than male students, and their answers elicited relatively shallow responses from their teachers (Sterling, 2000). President Freeman Hrabowski of the University of Maryland Baltimore Country (UMBC) gave an address on the Meyerhoff Scholars program at UMBC and the remarkable success UMBC has had in educating African American males (New York Times Editorial Board, 2013) who then often reach the highest levels of excellence in science. Coalition member, former astronaut, and University of Virginia Professor of Engineering, Kathryn Thornton, spoke about conflicts and opportunities faced by women scientists who are also wives and mothers (Thornton, 2000).

As a part of the conference the Coalition recognized 11 Programs That Work in an awards ceremony officiated by Virginia Secretary of Education, Wilbert Bryant. Leaders of each of the programs made a presentation at the conference. Volume 3 Issue 2 of the Journal of Mathematics and Science: Collaborative Explorations was devoted to the conference. The special issue consisted of papers describing each of the recognized programs, as well as the papers by Thorton and Sterling that are described above.

The conference was kicked off by Coalition President, Mark Warner, and Coalition Director, Loren Pitt, giving the keynote address entitled Challenging the Coalition: Raising the Participation and Success of Women and Minorities in Mathematics, Science, and Technology. Warner and Pitt (2000) stated: “Members of the VMSC strongly believe that many more students can be successful in mathematics and science. The question is how to go forward” (p. 2). The journal special issue concluded with a set of recommendations from conference participants. The
final recommendation was that the Coalition should continue Programs That Work as a way “to encourage the individual, personalized, caring approaches that can make the difference in individual lives” (Virginia Mathematics and Science Coalition, 2000, p. 111).

Programs that Work becomes an Annual Coalition Tradition

The Coalition followed this recommendation and has continued to recognize effective STEM programs in Virginia through Programs That Work. Dominion Power has also continued to provide financial support. Programs were recognized and award ceremonies conducted every year until the COVID-19 pandemic necessitated a two-year suspension. Programs That Work resumed in fall 2023.

Recognizing K-12 Public Schools

Over the two decades since Programs That Work was launched, Virginia faculty and leaders at all levels have been recognized for developing and offering effective STEM focused programs. Programs that Work recognizes programs offered to K-12 and undergraduate students as well as professional development opportunities to provide new knowledge and experience for in-service and preservice teachers. Many of the programs were designed by K-12 teachers and district leaders for students in their schools. For example, the Stafford County Public School district was recognized for providing meaningful watershed experiences for middle school students. The Frederick County Public School district was recognized for using carpentry activities to develop and enhance the mathematics skills of high school students. Henrico County Public Schools developed engineering experiences for elementary school students to introduce them to STEM fields. Granby High School in Norfolk was recognized for its Girls in Engineering after school program which included field trips to engineering-focused corporations, robotics classes that led to entries in an international robotics competition, and presentations by successful women scientists.

Other school systems were recognized for providing effective professional development for their teachers. The Comprehensive Instructional Project was initiated in 2014 as a consortium of 30 public school divisions in Virginia’s Region VII. The most successful teachers who taught some of the most at-risk students were recruited from the highest performing school divisions based on Virginia Standards of Learning test scores. These individuals prepared pacing guides, lesson plans, instructional materials, assessments and benchmark assessments as resources for the teachers in the region. In another professional development program, Arlington Public Schools was recognized for the effective mathematics content for elementary mathematics teachers that was developed and offered by mathematics specialists.

The Loudon County Public School district was recognized for after school STEM programs designed to help economically vulnerable students prepare for and enroll in the district’s gifted program. These 4th through 7th grade students who had a high score on one, but not both, of the cognitive ability tests (Naglieri or COGAT) that are administered as entrance exams were ineligible for the gifted program. The after-school program was designed to increase the problem-solving skills and collaboration skills of the students, identify their gifted behaviors, and to then have them referred to the gifted program by district faculty or administrators.
Recognizing Institutions of Higher Education

Institutes of higher education were recognized through Programs that Work for offering STEM-focused opportunities for their students. One of these programs was led by author Deborah Neely-Fisher of J. Sargeant Reynolds Community College (Reynolds) who, at the time, was not yet a member of the Coalition. The program included a research apprenticeship offered by Reynolds and Virginia Commonwealth University (VCU) to encourage and inspire minority students to major in disciplines that led to careers in biomedical research. Hampden-Sydney College was recognized for developing a laboratory module that promoted undergraduate understanding of gene expression and introduced students to qualitative and quantitative methods for studying this process at both the single-cell and population levels. The Department of Geology and Environmental Science at James Madison University was recognized for developing an effective Bachelor of Arts in Earth Science. The program is based on a holistic approach to the study of the earth by integrating all subjects including oceanography, meteorology, and astronomy into the undergraduate degree program.

Universities were also recognized for providing support for in-service teachers. At VCU, author Bill Haver and others were recognized for their role in developing the Virginia Mathematics Specialist Program that prepares experienced teachers to serve as mathematics specialists in Virginia’s school districts. Mathematics specialists provide in-school support for elementary and middle school teachers. Research conducted by project principals demonstrated the impact that mathematics specialists have on student achievement. Sweet Briar College was recognized for a professional development program introducing in-service teachers to using inquiry-based learning methods to teach STEM courses. George Mason University (GMU) was recognized for developing a professional development program as well as coaching and mentoring opportunities to support new secondary school science teachers.

Recognizing Other Institutions

In addition to K-16 educational institutions, other groups and organizations have been recognized through Programs that Work. The Virginia Department of Game and Inland Fisheries was recognized for providing in-service training and assistance to teachers and non-formal educators through an exemplary wildlife curriculum using hands-on activities to support student development of the Ecology Standards of Learning, the Virginia Environmental Science Concepts, and the Virginia Foundation Blocks for four-year-olds. It provides support for statewide facilitators on using the interdisciplinary conservation and environmental activity guides developed by the Association of Fish and Wildlife Agencies including Project WILD, Aquatic WILD, Flying WILD, and Growing Up WILD.

The Center for Excellence in Education in McLean, Virginia aims to empower underserved and underprepared students to pursue careers of excellence and leadership in STEM. They work with community colleges, universities, state agencies, and fellow STEM organizations to collaboratively improve STEM teaching and learning. The organization was recognized for providing opportunities for middle and high school teachers in underserved schools to connect with leading experts in industry and academia, to explore cutting-edge research, and to make other types of meaningful professional connections that have direct benefits for their students.
Outstanding Hosts and Keynote Speakers

The programs mentioned above are just a sample of the many outstanding programs that have been recognized during the last 23 years. Each year, the Coalition and Dominion Energy organize a recognition ceremony that enables participants to learn about the recognized projects through poster sessions or presentations by that year’s awardees. The events have been hosted by a variety of institutions including the Virginia General Assembly, the Science Museum of Virginia, VCU, the University of Virginia, and Reynolds. In addition to the awardees and the members of the Coalition, the ceremonies have been attended by business leaders, state office holders, school and university faculty and administrators. Legislators have also frequently attended, especially through the initiative of Coalition past president, John Watkins. Keynote addresses have been given at the ceremony by Secretary of Education, Anne Holton; legislator and Coalition president, John Watkins; Director of the Mathematics and Science Center, Julia Cothron; lawyer and Coalition member, Speaker Pollard; Science Museum of Virginia chief scientist, Jeremy Hoffman; business leaders; and researchers who have shared strategies for effective education practices.

Interested readers can peruse many of the Programs that Work awardees by visiting https://www.vamsc.org/index.php/programs-that-work/.

Programs that Work Resumes

A recognition ceremony was held in February 2020, just before the COVID-19 pandemic necessitated the cancellation of many public events. After that, Programs That Work was in hiatus until fall 2023 when a call for nominations was made by the Coalition. Deborah Neely-Fisher led the overall activity, handling many logistical aspects, with support from Bill Haver. Through the review process seven programs were identified as worthy of recognition. Each of the recognized programs actively engaged students and teachers. Most of the programs involved hands-on, project-based learning activities. Many involved computational thinking and data handling. Each program has a multi-year successful track record. The Poster Session and Awards Ceremony took place at Reynolds on November 3, 2023. Reynolds donated the use of its conference center for the ceremony and Dominion Energy provided funding for Programs That Work with a donation, as it has since 2000.

Participants were welcomed by Lori Dwyer, Vice President of Academic Affairs at Reynolds. Anne Peterson, Science Coordinator from the Virginia Department of Education (VDOE), brought greetings from the VDOE. It is interesting to note that ten years earlier, long before she joined VDOE, Programs that Work recognized Anne’s efforts to develop and implement a STEM team program for high school students in Gloucester County.

Representatives of each recognized program introduced their poster and attendees then engaged in animated discussions at each of the posters which were spaced around the conference center. Michael Broda and Sharon Zumbrunn from the VCU School of Education presented the keynote address entitled *Calculated Success: The Power and Potential Pitfalls of Growth Mindset and Belonging in STEM*. They described situations in which minority students do not have a sense of belonging which, in turn, impacts their educational success. This was reminiscent of the research that Donna Sterling reported during the first Programs That Work conference in 2000 bringing the ideas full-circle that female students were called upon much less frequently than are males, and their answers elicit relatively shallow responses from their teachers (Sterling, 2000). Plaques
were presented to each recognized program by Coalition members Deborah Neely-Fisher, Elizabeth Edmondson, and Speaker Pollard. A summary of each recognized program is provided below.

**Programs that Work Recipients 2023**

**Newport News Public Schools STEAM Camps:** Tami Byron, Newport News Public Schools (NNPS). This program consists of 3-week summer camps designed for students in grades 3 through 12. In 2023, a total of 10 summer camps were offered. The camps provided hands-on learning experiences in computer science, emerging technologies, and engineering design. They were designed to spark creativity in students, promote teamwork, and teach real-world problem-solving skills. To further enrich the experience the students were mentored by William & Mary preservice teachers and STEM majors and by NNPS juniors and seniors. The STEAM Camps have proven to be very popular with students. The camps were initiated with a Department of Defense grant in 2019 serving 60 students; in 2023, 700 hundred students applied and, through the use of a lottery system, 350 students participated in the camps. Students also hosted a Learning Expo for the community, providing a practical assessment of their work. Pre- and post-engagement surveys demonstrated increased engagement by participating students in all STEM disciplines.

**Solar Baby: Allowing Students During Enrichment Time to Compete in a Solar Challenge:** Tracy Rhodes, J. Frank Hillyard Middle School, Rockingham County Public Schools. Students are given the opportunity to use enrichment time to participate in a project-based learning activity. The students learn about solar energy, renewable energy resources, circuits, coding, collaboration, presentation skills, science and engineering practices, gear ratios, and load capacities. Teams of students construct and test their projects: wind turbines, solar boats, solar towns, or other solar structures. The program is in its third year; during the 2022–2023 school year 21 students participated. The solar teams spent a morning at a local elementary school with fifth graders describing how the circuits and solar panels worked on their projects and engaged the fifth graders in completing circuits to run an LED bulb. A central activity for each team was entering their project in the KidWind Solar Challenge competition beginning at the school and with the potential to compete at the regional, state, or national level. After each competition students use the feedback from the judges and modify their design. During the 2022–2023 academic year, two teams qualified for the state competition and one team qualified for the national competition. The two teams that participated in the state competition placed second and third. The team that went on to the nationals placed first in the nation.

**High School Student-Led Research Around Data Science and Computing:** Shenandoah Valley Computer Science Regional Partnership; Deb Crawford, Frederick County Public Schools (FCPS); and Padhu Seshaiyer, GMU. The Shenandoah Valley Computer Science Regional Partnership developed and offered a virtual high school student-led research program. Under this program, that has been offered for the last two years, 30 students engaged in data-focused research using multiple disciplines combining mathematics, statistics, and computer science skills to solve societal challenges. The students each identified a potential challenge that aligned with one of the 14 global engineering challenges addressing the needs of the poor, the environment, or a future challenge for improving life on the planet as identified by the National Academy of Engineering. The students conducted research on the challenge guided by a faculty mentor through weekly virtual meetings. At the conclusion of the program, each student
showcased their research for the students, teachers, and parents from the partnering districts. They earned two dual enrollment credits through GMU. The student-led research program was well received and has led to the initial phases of development for a lab school, the regular offering of a formally approved GMU course and the creation of the first Virginia Data Science Standards.

**Designing the Science Practices Innovation Notebook (SPIN):** Dr. Erin Peters-Burton, GMU. This three-year professional development program was offered to 20 Loudon County teachers from schools serving low socio-economic status students and from schools serving high achieving students. The program focused on having the teachers, in partnership with program leaders, develop and refine a web-based science notebook, SPIN, that presents science investigations on data practices, supports computational thinking, and encourages students’ self-regulated learning through a series of customizable prompts. SPIN was developed over a three-year period during which the different components of the materials were repeatedly field tested. Evidence demonstrated that teachers continually achieved learning gains over the three-year timespan, even when faced with challenges associated with the COVID-19 pandemic. And, across the years, teacher learning became progressively more sophisticated, complex, and contextualized. To a large degree, the participating teachers persisted in the long-term professional development opportunity because they were motivated to produce SPIN for use by other teachers.

**Enhancing STEM Engagement Through Solar Car Builds with the Flying Classroom:** Decardra Jackson, Petersburg Public Schools. Middle school students engaged in a long-term project to explore solar energy, electricity, and sustainable energy sources through an automotive lens. The project combined hands-on learning, renewable energy education, and exposure to advanced careers in the automotive industry. Students studied STEM topics such as lengths, proportions, angles, and ratios and then made use of the concepts through hands-on experiences in welding and working with steel, and by conducting data analysis on vehicle diagnostics and control systems. Carrying out their project also allowed the students to learn about photovoltaic cells, energy conversion and harnessing sunlight to produce electricity. Quantitative pre- and post-assessments and qualitative teacher observations, student reflections and peer assessments demonstrated that participants had acquired knowledge related to technology, renewable energy sources, electrical systems, and mathematics calculations, and that they have developed critical thinking, problem-solving, and decision-making skills.

**Shenandoah Valley Computer Science Regional Partnership:** Padhu Seshaiyer, GMU and Deb Crawford, FCPS. The Shenandoah Valley Computer Science Regional Partnership consists of seven school systems, the regional governor’s school and higher educational institutions. The partnership was formed in 2019 to increase opportunities for K – 12 students in the Valley in the areas of K – 8 Computational Thinking and high school Computer Science. The partnership has obtained support to provide extensive professional development for teachers. For example, a K – 5 virtual webinar series on implementing the Computer Science Standards of Learning served about 275 teachers and coaches in the region. A computer science teacher leader now serves as a coach in each elementary school in the partnership; 37 Grade 4 – 8 language arts and English learners teachers and coaches were trained in using Twine, an open-source tool for telling interactive stories, to improve writing across the curriculum through storytelling. An 18-credit 100% online program for teachers to earn a computer science add-on endorsement was developed and 53 teachers are currently earning credits through the program. Through these and
other teacher initiatives, integrated approaches to learning of computer technology and computer science are being institutionalized in the partner schools at all grade levels.

**Southeastern Virginia Environmental Education Consortium:** Dr. Venicia Ferrell, The Center for Educational Partnerships at Old Dominion University. The constituent divisions of the Southeastern Virginia Environmental Education (SEVEE) Consortium are Hampton City Schools, NNPS, Norfolk City Schools, Portsmouth City Schools, Suffolk Schools, and Williamsburg-James City County Schools. Eighty percent of the participating schools deal with environmental justice inequities. Those schools are in Virginia municipalities with high numbers of low-income families, communities of color, low education levels, and linguistically isolated, non-English speaking households. These schools have high EPA environmental justice indices reflecting that low-income and minority residents are at risk for environmental hazards and are close to areas experiencing high wastewater discharge, posing potential risks to nearby rivers, streams, and the Chesapeake Bay. Through the consortium, 67 teacher leaders from constituent schools were prepared to support the systematic implementation of Meaningful Watershed Education Experiences (MWEEs), based on the National Oceanic and Atmospheric Association framework, within their schools. Lessons providing MWEEs for elementary, middle, and high school students were developed and shared through the SEVEE website. As indicated on post program surveys, teachers valued the professional development opportunity with over 92% rating the overall professional development experience highly and 93% stating they were impressed with the quality of examples and lessons on watershed and the Chesapeake Bay. In addition to the expert-led professional development activities, teachers had the opportunity to share their in-class environmental literacy projects. Teachers and students discussed their field experiences and MWEEs with community partners. Model research and service projects are being shared across divisions; “Pop Up” professional development opportunities have occurred across each school division. This effort has been fundamental in creating an environmental-focused professional social network across the region.

**Virginia Educators Deliver High Quality STEM Programming**

The longevity of Programs That Work is testimony to the health and vitality of STEM educational programs across Virginia. Faculty respond to the need to improve the environment and to environmental justice inequities by creating programs like those developed by the Southeastern Virginia Environmental Education (SEVEE) Consortium. Business leaders have called for schools to produce students with strong computer skills. The Shenandoah Valley Computer Science Regional Partnership is responding to that need. There is a national call for renewable energy and, as a result, the national KidWind Solar Challenge has been created. Faculty at Rockingham County Public Schools also responded to this situation by engaging their students to better understand solar power through a project-based learning activity. Programs were developed in Petersburg Schools to engage minority students in educational programs that incorporate STEM topics with hands-on experiences in welding, working with steel, and conducting data analysis on vehicle diagnostics. A large-scale summer camp program has been developed and offered for students in NNPS. And Loudon County and GMU developed a new approach to professional development through the creation of a web-based science notebook in a process that was able to remain vibrant during a pandemic.

The Coalition appreciates these evolving challenges as well as the opportunities they presented for institutions and organizations in Virginia. Through Programs That Work, the
Coalition both recognizes the creativity and achievements of STEM faculty and enables others to learn from and benefit from these effective approaches.

References


Virginia Mathematics and Science Coalition. (2000). Recommendations from conference participants. *Journal of Mathematics and Science: Collaborative Explorations*, 3(2), 107 – 111. [https://doi.org/10.25891/ej0g-j388](https://doi.org/10.25891/ej0g-j388)

INTRODUCING FLEXIBLE ASSESSMENT INTO A COMPUTER NETWORKS COURSE: A CASE STUDY

Joseph Meehean  
University of Lynchburg  
meehean.j@lynchburg.edu

ABSTRACT
With overall positive results and limited drawbacks, I have adapted modern pedagogical techniques to address a common difficulty encountered when teaching a computer networks course. Due to the tiered nature of the skills taught in the course, students often fail unnecessarily. Using mastery learning, competency-based education, and specifications grading as a foundation, I have developed a course that allows students with varied skills and abilities to pass. The heart of this approach is the flexible assessment of programming assignments which eliminates due dates and allows students to have their work graded and regraded without penalty. Flexible assessment also defines an interactive approach to grading which gives students immediate formative feedback and does not penalize initial failure. Using these instructional techniques, I improved the course completion rate by 30 percentage points compared to similar courses. Flexible assessment works best for upper-level courses that are not prerequisite courses because a student can pass without mastering all of the skills; their grade reflects the percentage of skills mastered rather than an average of the competency of all the skills taught. Drawbacks of flexible assessment include limited time for in-class preparation, limited opportunities to review programming assignments and the increase in time required for grading.

KEYWORDS
mastering learning, specifications grading, flexible assessment, computer science education

https://doi.org/10.25891/en8j-cz52
A common challenge in creating a Computer Networks course is designing the course in a way that allows students with a variety of skill-levels and academic abilities to succeed. The tiered nature of the skills developed in the course means that students who stumble early are often unable to catch up. Lower performing students may master only a subset of the skills covered by an assignment, having only a limited proficiency of the remaining skills. Because each assignment builds on the full set of skills learned in previous assignments, a student who does not master an assignment starts the next assignment at a severe disadvantage. The cumulative nature of this disadvantage causes poor performing students to quickly fall so far behind that they cannot pass the class.

Borrowing techniques from competency-based learning (Gervais, 2016), specification grading (Berns, 2020; Nilson, 2014; Sanft et al., 2021) and mastery learning (Bloom, 1968; Garner et al., 2019; Keller, 1968), I have developed a Computer Networks course that allows lower performing students to pass. Programming assignments are often the component that prevents struggling students from passing the class. In response, I have created a flexible assessment pedagogy to allow the students to develop skills at their own pace. In this pedagogy, there are no due dates for assignments and I regrade assignments without penalty as often as the student likes. Students schedule an interactive, formative grading session with me whenever they feel their project is finished or has seen substantial improvement since the last grading session. This style of grading gives students immediate feedback and allows them to develop and build confidence.

Early results indicate that the flexible assessment pedagogy shows promise. It has increased the number of students who finish the course with a passing grade by 30 percentage points when compared to courses with a similar difficulty and audience.

There are some drawbacks and limitations to this technique. The flexible assessment pedagogy will not work for courses that are prerequisites for other courses because a student may not learn all of the prerequisite skills. Any skills they do learn, however, they master. The lack of due dates means that different students are often working on different programming assignments at the same time. This limits the amount of in-class review I am able to do for programming assignments. Perhaps the biggest drawback for the instructor is the amount of time spent grading.

Related Work

I have not developed radical new teaching methods. Rather, I have borrowed and adapted existing instructional approaches to fit the needs of my Computer Networks course and my students. The biggest influences for flexible assessment are competency-based education (Gervais, 2016), mastery learning (Bloom, 1968; Garner et al., 2019; Keller, 1968), and specifications grading (Berns, 2020; Nilson, 2014; Sanft et al., 2021).

Competency-based education (Gervais, 2016) focuses on students developing a predetermined set of competencies or real-world skills. The term is broadly used, but some hallmarks include unlimited regrading, self-pacing, student-directed learning, and frequent formative assessment. The largest difference between my Computer Networks course and a traditional competency-based course is that this course was not student-directed. I provided a schedule for course material, gave lectures, and assigned written homework that covered the theoretical aspects of the course content. Another major difference is that students were given 15 weeks to complete as many assignments as possible, unlike a traditional competency-based course where students have as much time as they like.
Mastery learning proponents posit that 90% of students can master the required learning outcomes of a course given sufficient time and help from the instructor. Common themes in mastery learning are that students work at their own pace, participate in formative evaluations with detailed feedback, and repeat assessments until they demonstrate mastery and, therefore, are ready to move on to the next topic. Like Bloom’s (1968) Learning for Mastery and Keller’s (1968) Personalized System of Instruction, my form of flexible assessment allows students to work at their own pace and provides formative evaluations with detailed feedback. In contrast to Bloom and Keller’s models, my form of flexible assessment does not require students to achieve full mastery of one programming assignment before moving on to the next. Additionally, my course is time-limited, so I do not expect that 90% of the students will achieve mastery of all the required skills. Garner et al. (2019) provide a literature review of the use of mastery-based learning in computer science courses. The review includes only a small number of papers, which limits their conclusions. The authors conclude that most universities using mastery-based learning do so only for introductory computer science courses and that instructors often add traditional exams to provide summative assessment. The review finds that the primary motivation for transitioning to mastery learning is to improve learning outcomes in student populations with wide-ranging aptitudes.

Specifications grading (Nilson, 2014) creates a tiered assessment system. Each assignment grade is binary, pass or fail. Assignments are grouped into bundles or modules. A certain group of bundles must be passed before a student can earn a specific grade (e.g., bundles 1 – 2 for a D, bundles 1 – 4 for a C, etc.). This model includes a limited number of resubmissions for failed assignments. Modified specifications grading has been applied by other computer science instructors. Berns (2020) proposes a system called binary grading which is similar to specifications grading but allows unlimited resubmissions. Sanft et al. (2021) propose a modified specifications grading system with pass/fail assignments. Resubmissions are allowed, but incur a 10% penalty per retry. Sanft et al. show an improvement in student learning outcomes for middle to low performing students. The flexible assessment pedagogy that I developed allows both unlimited resubmissions and partial credit, rather than pass/fail, for programming assignments.

Concurrently with my own work, Lionelle et al. (2023) have developed their own flexible assessment model with a focus on large introductory computer science courses. This model includes formative assignments with no deadlines and no-penalty resubmissions. This model also adds a small number of summative assignments with a limited number of resubmissions and hard deadlines. Mastery of a topic is required for students to move to the next. This model has shown improved student performance in courses that follow the introductory courses, demonstrating that students are mastering foundational skills. In contrast to my own flexible assessment pedagogy, Lionelle et al. make extensive use of auto grading systems which limits the type of feedback that can be provided to students. This limitation is a result of focusing on large classes where one-on-one formative assessments would be impossible.

Problem Addressed and Desired Aspects of a Solution

Students often fail Computer Networks unnecessarily. The skills and knowledge required for programming assignments in the course build on one another. That is, a student who does not successfully complete the first assignment cannot complete the second assignment since the skills developed in the first assignment are a prerequisite. Using a traditional approach to
teaching this course, a student who stumbles on the first assignment may never catch up and will ultimately fail the course.

With a perfect solution to this problem, a student who masters a subset of the skills taught in this class would pass. Students could learn a subset of skills at their own pace. And, any skill a student does learn, they learn well. Students who master all of the skills earn an A; students who master fewer skills earn a lower grade.

Course Description

Computer Networks is the study of the design and use of computer networks, with a focus on the modern Internet. The course focuses on the theoretical underpinnings of the modern Internet and the specific algorithms used to implement it. In particular, this course discusses client-server programming and its relation to the application, transport, network, data, and physical layer protocols of the Internet. Computer Networks is an upper-level elective for computer science majors and minors at the University of Lynchburg. Roughly half of our upper-level students opt to take this course. The content of this course is divided into theory and practice. The theory portion is focused on the foundational ideas and algorithms that define the modern Internet. The practical component, on the other hand, focuses on learning how to use the powerful network interfaces provided by programming languages, software libraries, and operating systems. Most graduates who work in computer networks will be developing software which uses the network, a practical application of this course. As good professionals, they should have a strong grasp of the theoretical foundations that power the networks they are using. Therefore, class time is divided between lectures, which cover the theoretical aspects, and hands-on-activities that cover the practical aspects. Students work in small groups to discuss and understand the theoretical concepts and they may work in pairs on the practical components.

Course Setup

Assignments for this course are divided into two categories: written assignments and programming assignments. Written assignments cover the theoretical aspects of computer network design and are intended to replace traditional exams. There are five of these assignments and they consist of four to five short answer questions. Students are given a week to complete each of these assignments. The typical student can complete them in a few hours. These assignments align with a traditional form of assessment: they have a fixed start and due date and students are only allowed to attempt them once, without regrades. These assignments compose 35% of the final grade.

The programming assignments provide students with practice writing software that use computer networks. There are three programming assignments and the typical student needs several weeks consisting of nine to ten hours per week of work to complete them. These assignments are the heart of the flexible assessment pedagogy. There are no fixed due dates except that they have to be completed by the end of the semester. Students are given freedom to decide how they complete these assignments. I also provide flexibility in the grading of these assignments; students can have these assignments graded as often as they like without penalty. These assignments comprise 65% of the final grade.

Through these programming assignments students build a simple client-server network file system, like a greatly simplified File Transfer Protocol (FTP) (Postel & Reynolds, 1985).
Each assignment builds upon the skills (and sometimes the code) students developed in the previous assignment. The first assignment focuses on building a network-based key-value store (see Appendix A for the full assignment). A key-value store is a simple database that stores key-value pairs. A user can ask a key-value store to save a key like “email-address” associated with a value like “somebody@example.edu”. Later the user can ask the key-value store to retrieve the value for the key (email-address) to get back the associated value (somebody@example.edu).

For students, the difficult part of this project is developing the network protocol to allow the data to be stored on one computer and accessed or updated from another computer. In the second assignment, students convert their key-value store into a network flat file system which is a file system that does not have folders or directories. At the end of the second assignment students have a network-based file system that allowed files to be added, retrieved, and appended to. The third and final assignment improves the reliability of the flat file system by adding features that could gracefully handle crashed or frozen clients and servers, protocol errors, or corrupted files.

I created these assignments thematically to allow students to demonstrate a mastery of a collection of skills that cover aspects of network programming. Program 1 contains a set of skills a programmer needs to build a rudimentary, text-based, client-server network application. Program 2 introduces the skills required to build a more complex client-server network application that can handle non-text data like video files. And Program 3 includes techniques to make a network application resilient to failures. As an example, see Table 1 for the set of skills developed by a programmer who completes Program 1.

Table 1
Skills Demonstrated by Completing Program 1

<table>
<thead>
<tr>
<th>Program 1 Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiate and accept TCP connections</td>
</tr>
<tr>
<td>Send and respond to text-based commands</td>
</tr>
<tr>
<td>Develop a protocol to differentiate between commands and user data</td>
</tr>
<tr>
<td>Develop a protocol to differentiate between text-based errors and requested data</td>
</tr>
<tr>
<td>Develop a protocol that allows all text-based data to be transmitted (incl. special characters)</td>
</tr>
<tr>
<td>Gracefully recover from non-fatal errors in the server</td>
</tr>
</tbody>
</table>

Several aspects of these programming assignments are more flexible than their counterparts in a traditional networks course. Students are allowed to develop their own network protocol to govern communication between the client software and the file server. A network protocol defines the ordering of network messages (e.g., who initiates the communication) and the data format of the messages. In a more traditional networks course, the professor would define this protocol for the students. Students write the code for these assignments using the Java programming language. Java provides a myriad of ways to interface with the computer network, from low-level interactions using bytes to sending high-level programmer defined objects. I give the students the freedom to use any of Java’s network interfacing tools. Again, in a traditional class, students would be given specific directions about how to use Java in their assignments.

The most flexible aspect of this approach is the lack of due dates or a rigid sequencing of the assignments. When a student feels that their project is ready, I grade it. Students earn points for each feature of the assignment that works to specification. If during the grading session a
feature does not work, the student is encouraged to spend more time working on it. When the student fixes the broken feature, I regrade the assignment. The student receives full points for each additional feature that works. This iterative approach to grading gives the student multiple opportunities to learn a skill with helpful feedback along the way. When a student feels that they have learned enough from an assignment, they move on to the next assignment, even if they have not developed all of the features of the previous one. This is a significant contrast to mastery learning and specifications grading. If the next assignment proves to be too difficult, the student can return to a previous assignment to reinforce the skills that it covered and earn more points. This is in stark contrast to the traditional rigid sequencing of a computer science course where a student is given a fixed amount of time to complete an assignment and must move on to the next assignment whether they are ready or not. The flexible sequencing of assignments in this course matches the nature of network software development quite well. A student must understand the rudimentary aspects of network protocols and network software development before they can attempt to build more complex real-world network applications. In a traditional networks course, if a student struggles with the first programming assignment it is impossible for them to complete any of the following assignments. This ensures that they fail the class. Using the flexible approach, students accrue network development skills at their own pace.

A key part of this flexible approach is setting clear expectations and standards for the students. At the start of each assignment, students are given the grading rubric (see Appendix B for the student’s version of the rubric for Program 1). The rubric format is a list of three types of features: prerequisite required features, features that earn points along with their point values, and features that are not required. Students need to have all of the prerequisites completed before I will grade the assignment. These prerequisites are generally straight-forward and ensure that students do not violate the spirit of the assignment (e.g., the project must use a computer network). For each of the features that earn points, the rubric specifies how many points each feature is worth and provides general examples of the specific behaviors that must be supported (e.g., server should reject keys that are already in the key-value store). I include a list of features that are not required to prevent students from wasting valuable time building things that are beyond the scope of the project (e.g., the client does not need to have a graphical user interface). These rubrics are designed to help students focus on the most important aspects of each assignment and prevent surprises during grading.

On a broader scale, I also have a grading rubric for the full set of programming assignments. Students who earn all of the points on the first assignment earn a passing grade for the programming assignment portion of the class. Students who also earn all of the points on the second assignment earn a B –, and students who earn all of the points for all three assignments earn an A. I created this overall rubric to motivate students to complete all three assignments and to prevent them from being surprised by their final grades.

**During the Course**

During the course, the two most important parts of the flexible assessment process are student preparation and grading.
Student Preparation

Preparation for the first programming assignment is vitally important because the lack of due dates means I cannot host a post-assignment review session after the due date to discuss the proper way to complete the assignment.

After the first five weeks, most of the material the students need to complete the first programming assignment has been covered through the course lectures. In addition to the lectures, I use a class period to host a code-along during which we built a simple echo server and a National Institutes of Standards and Technology (NIST) time server client (Lombardi, 2002). A code-along is an interactive, in-class activity where the students and instructor write software together. During the code-along, I explain the feature we are trying to build and the Java classes we will use to build it. I then give them some time to try to build a NIST time server client on their own. After they spend some time working on it, we come back together as a whole group and they explain to me how to write the code while I type. With a little guidance from me, we are able to get a specific feature of the program working. Then we move on to the next feature and repeat the process. In this way, each student has two working network programs and several weeks of lectures on which to base the development of their first programming assignment.

Flexible Grading

The grading process is fairly straightforward and starts with the student. When a student feels that their assignment is ready to be graded, they schedule an individual meeting with me. During the meeting the student demonstrates the working features of their assignment. For each assignment there are specific test cases I use; these test cases cover aspects of the assignment students may not have considered. I record which parts of their assignment work to specification and which do not. Initially this grading process was slow for me, but by the end of the semester I had it down to about 10 to 15 minutes per student per assignment.

When a student’s program correctly implements all of the features listed in the rubric, they have demonstrated mastery of a collection of related network programming concepts. In this way, a student that earns a C has demonstrated mastery of a subset of the skills required for this course. This approach to assessment differs from a traditional course where a C may mean the same as above or it may mean that the student only partially understands all of the course material.

In a traditional course, grading is often a form of summative feedback and can be viewed as pure assessment. Through the flexible assessment approach, grading gives students formative feedback that helps them better understand the skills they are learning. Students can use this feedback to improve their project and earn more points. This incentivizes quick integration of feedback to improve their skills. Another important feature is the interactive nature of the feedback. If a feature does not work to specification, we discuss why. Students sometimes misunderstand a core concept from the lectures and I take the time to re-explain it to them. Or they simply misunderstand the feature I am asking them to build. Summative feedback penalizes these misunderstandings. The formative feedback allows students to expose and repair gaps in their knowledge without fear of it effecting their grade.

Without due dates, one of my largest concerns is academic misconduct. I do not want one student to finish the assignment and immediately give the answers to everyone else. To prevent
this, each grading session starts out with a conversation. The student explains the network protocol they developed for the assignment and answers some questions about the project. To prevent the student from feeling tricked, I give them these questions in advance. If the student cannot explain their protocol or answer the questions satisfactorily, we stop the grading session, no points are awarded, and I give them advice on how to prepare for the next grading session. This approach prevents students from sharing too much information about their assignments with their classmates. Students can give each other limited help, but they know that the help is only the start. Each student needs to cultivate a deeper understanding of why the project is built a certain way or they will not be able to answer my questions. Since this deeper understanding is a goal for this course, in my opinion, how it is acquired is somewhat irrelevant.

Results

During my first semester of using flexible assessment, the vast majority of students acquired some computer networking skills during this course. As shown in Table 2, by the end of the semester, 80% of students were able to demonstrate a complete collection of skills required to implement a simple computer networking program. And over half were able to acquire at least one skill at an advanced level. While just one student was able to demonstrate competency of all of the skills at the advanced level.

Table 2
Percent of Students whose Programs Met the Required Specifications

<table>
<thead>
<tr>
<th>Programming Assignment</th>
<th>Fully met the specifications</th>
<th>Met 80% of the specifications</th>
<th>At least one feature met the specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program 1: Key-value store</td>
<td>80%</td>
<td>86%</td>
<td>93%</td>
</tr>
<tr>
<td>Program 2: Flat File System</td>
<td>60%</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Program 3: Reliable File System</td>
<td>6%</td>
<td>20%</td>
<td>53%</td>
</tr>
</tbody>
</table>

It is difficult for me to evaluate the success of this course compared to a more traditional approach because this was the first time that I taught Computer Networks and my university only offers one section of the course each academic year. As a result, I do not have data from a traditional networks class for comparison. However, I also teach the Distributed Systems and Operating Systems courses which are of a similar difficulty and have a similar audience. So I compared student performance in Computer Networks to student performance in my previous two sections of Distributed Systems and my previous two sections of Operating Systems. Distributed Systems uses a similar interactive assessment style, but with rigid due dates and without unlimited regrades. Operating Systems uses a traditional time-restricted, summative assessment system where students submit their work and I grade it without them being present.

In order to compare these courses, I defined two metrics: completion percent and passing percent. I defined completion percent to be the percentage of students who finished the course with a grade better than an F divided by the number of students who started the course. The passing percent is the percentage of students who finished the course with a grade better than an F divided by the number of students who did not withdraw from the course. In should be noted
that at my university, students who are struggling are allowed to withdraw during the first two-thirds of the semester, so passing percent should be higher than the completion percent.

Table 3
Pass, Fail, and Withdraw Data for Three Similar Computer Science Courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Students</th>
<th>Pass</th>
<th>Fail</th>
<th>Withdraw</th>
<th>Completion %</th>
<th>Passing %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributed Systems*</td>
<td>29</td>
<td>19</td>
<td>4</td>
<td>6</td>
<td>65%</td>
<td>82%</td>
</tr>
<tr>
<td>Operating Systems*</td>
<td>30</td>
<td>14</td>
<td>1</td>
<td>15</td>
<td>47%</td>
<td>93%</td>
</tr>
<tr>
<td>Computer Networks</td>
<td>15</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>87%</td>
<td>93%</td>
</tr>
</tbody>
</table>

* Two sections combined.

Table 3 shows the student completion percent for Computer Networks with flexible assessment was much higher than the rates for the comparable courses, but the passing rates were similar. Together, the comparable courses had a completion rate of 56% and a passing rate of 87%. Flexible assessment resulted in an increase of over 30 percentage points in the completion rate compared to the other two courses combined. This resulted in roughly five more students per section completing this course with a passing grade. The sample sizes here are, of course, small and it can be difficult to infer from these results whether future sections of the course will have similar results.

Table 4 shows the similarities and differences in grades across these three courses. Extracting meaningful patterns from this data is a bit more difficult. It appears that the flexible assessment approach is pushing students up from the bottom. It appears that withdrawals become Fs or Ds while Fs and Ds become Bs. This matches expectations from previous research which applied modified specification grading to computer science courses (Sanft et al., 2021). However, data from my course is likely too small to draw any conclusions about specific grades.

Table 4
Student Grade Distributions

<table>
<thead>
<tr>
<th>Course</th>
<th>Students</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributed Systems*</td>
<td>29</td>
<td>35%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>14%</td>
<td>21%</td>
</tr>
<tr>
<td>Operating Systems*</td>
<td>30</td>
<td>14%</td>
<td>27%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>50%</td>
</tr>
<tr>
<td>Computer Networks</td>
<td>15</td>
<td>32%</td>
<td>47%</td>
<td>0%</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
</tr>
</tbody>
</table>

* Two sections combined.

I also solicited student feedback on the flexible assessment aspect of Computer Networks. I asked students to evaluate different aspects of the flexible assessment features of the course. In particular, I asked if they found flexible assessment to be helpful to their learning process. They rated interactive grading, the lack of exams, and the lack of due dates on a Likert scale ranging from No Help to Great Help. The results appear in Table 5.

Students were somewhat split on the flexible assessment style of this course. Students liked not having exams and they had no complaints about the interactive grading. However, their feedback reflects that, overall, the class had a mixed view about the lack of due dates for the programming assignments. From their comments, it appears that some students felt that the absence of due dates caused them to procrastinate too much.
Table 5
Student Evaluation of How Much Each Aspect of the Course Helped Their Learning

<table>
<thead>
<tr>
<th>Flexible Assessment Feature</th>
<th>No Help</th>
<th>A Little Help</th>
<th>Moderate Help</th>
<th>Much Help</th>
<th>Great Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive Grading</td>
<td>0%</td>
<td>0%</td>
<td>14%</td>
<td>29%</td>
<td>57%</td>
</tr>
<tr>
<td>No exams</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>29%</td>
<td>71%</td>
</tr>
<tr>
<td>No due dates</td>
<td>14%</td>
<td>29%</td>
<td>14%</td>
<td>14%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Percentages are based on the number of students that selected each category.

These results are similar to other studies where modified specification grading was used in computer science courses (Berns, 2020; Santf et al., 2021). To this point I would note that procrastinating students are the ones who are most likely to fall behind on the first assignment in a traditional course and are, therefore, the students who are most likely to withdraw from a traditional course. Through this alternative approach, while procrastination is still painful, it is not academically fatal.

Along with this criticism came some student suggestions for improvement. They suggested requiring the first programming assignment be completed before the course withdrawal deadline. They also recommended putting suggested due dates on the assignments as a guide.

Difficulties and Limitations

This instructional approach is not without its difficulties and limitations. As stated above, the biggest limitation is that flexible assessment will not work for a course that is a prerequisite for another course.

A difficulty I had not considered when I decided to use flexible assessment for this course is the limits placed on in-class preparation for assignments and post-assignment review. The crux of this problem is that students are really spread out in their progress through the programming assignments: I had students who completed all of the assignments with weeks to spare and others who were working on the first assignment right up until the last day. In my other classes, I often provide code-along days to prepare for difficult assignments. I could only do that for the first programming assignment in this course. Doing a code-along for the second or third assignments would have given away too much information to the students still working on the earlier assignments. Similarly, in my other classes I often perform post-assignment reviews after an assignment has been submitted and graded. In these reviews, I discuss areas in which the class as a whole struggled and better approaches they could have taken. There was never a point during the semester where all my students had completed the first programming assignment, so I could not hold a review session without giving away too much information to students still working on the assignment.

An unsurprising difficulty with teaching a course using an interactive grading system with an unlimited number of regrades is that the grading takes a long time. Early in the semester, grading sessions were taking 30 minutes per student. The typical student scheduled three grading sessions for each assignment they completed. Later in the semester, I made modifications which reduced the time to 15 minutes; this still resulted in spending 45 minutes with each student per assignment over the course of the semester.
Another difficulty I had not considered was justifying to the students the use of the grading system. I naively assumed students would love it and so there would not be any concerns. Nearly all of the students accepted that the system was fair, even if they thought it had shortcomings. However, I did have one student who procrastinated to the point of failing the course. He was, of course, unhappy and so was his father. In a traditional course, I could point to a series of assignments with Fs that led to the final grade. With a different grading system, it can be hard to explain to a concerned person that the problem is not “this one assignment” but the fact that the student waited way too long to try to complete one assignment and, therefore, failed. Not to mention the student did not even attempt the other two assignments. So, this approach requires some thought on the professor’s part about how they will explain the grading process to students, parents, and administrators. It also requires support from the institution. If the administration does not encourage innovation or will not support faculty during a grade appeal, it may be better to stick to a more traditional assessment system.

Lessons Learned

As discussed earlier, grading takes a significant amount of time using the flexible assessment approach. During the course, I made some adjustments to speed up the grading process. First, I began to take and keep notes on each student’s project including how they defined their protocol and their work on the program features. This reduced the amount of time spent at the beginning of the grading session reviewing how the student’s project worked. I could simply ask what had changed since our last session. Second, I required a portion of the students’ assignments to meet a specific Application Programming Interface (API), a set of function or method signatures. For example, the client-side of the key value store needed to have a method called get that took a key as a parameter, sent the request to the server, and returned the server’s response. Prior to establishing a simple API, students had a myriad of ways of asking the client to retrieve a key’s value, several of which made grading slow. These required APIs sped up grading considerably and provided some additional structure for the students.

Reducing student procrastination is a much more difficult task. I have several improvements I intend to make to address this problem in future courses. Based on student suggestions, I will add recommended due dates to each of the programming assignments. Moreover, I intend to incentivize students to conduct their first grading session before the recommended completion date and the rubric for an assignment will award a few points for doing so. It would be tempting to view these points as a late penalty. I intend to discourage that in two ways. First the points will not be worth a full letter grade; they will be worth a half-letter grade, separating a B from B– for example. Secondly, the assignment does not need to work to specification by the suggested completion date to earn these points. The student only needs to sign up for and attend the first grading session to earn these points.

My final improvement for reducing student procrastination consists of modifications to the student suggestion that I require the first assignment to be completed by the withdrawal deadline. I want students to be able to pass this class, albeit with a D, even if it takes the entire semester for them to complete the first programming assignment. So, instead of requiring students to complete the first programming assignment by the withdrawal deadline, I will require students to attend at least one grading session and strongly encourage them to complete the first programming assignment before mid-semester. If the student’s first programming assignment is not graded prior to mid-semester, they will be withdrawn from the course for failing to make
satisfactory academic progress. The program does not need to work completely during the grading session and the student can schedule additional grading sessions for the assignment after mid-semester. However, if I grade their first programming assignment and it does not work to specification by mid-semester, the student will receive an F for their mid-semester grade, and I will email the student, their academic advisor, and the university’s advising department stating that the student is on track to fail this course. These tiered requirements should create a strong incentive for students to attempt to finish the first programming assignment prior to mid-semester. Getting students to start an assignment is often the hardest part, and I am hoping that once they start the assignment they will follow through and finish it. Additionally, if the student still has not completed the first programming assignment by the withdrawal deadline (i.e., about three weeks after mid-semester), I will recommend to the student and their academic advisor that the student withdraw from the class. This policy should also help mitigate the problem with students who claim to be surprised that they failed at the last moment because they did not finish “one assignment.” There will be a paper trail indicating that the student has known for weeks that they are in danger of failing.

Conclusion

Flexible assessment directly addresses a common difficulty encountered when teaching Computer Networks with overall positive results and limited drawbacks. I have adapted modern pedagogical techniques to improve the course which allows students with varied skills and abilities to pass. The heart of this approach is the flexible assessment of programming assignments which eliminates due dates and allows students to have their work graded and regraded without penalty. Flexible assessment also defines an interactive approach to grading which gives students immediate formative feedback and does not penalize initial failure. Using these techniques, I have increased the course completion percentage by 30 points when compared to similar courses. This approach also improved the learning outcomes for middle to low performing students. These results are similar to those found by researchers who modified specification grading for use in their computer science courses (Sanft et al., 2021). Overall, students liked the formative assessment style of this course, but some found the lack of deadlines to be an obstacle. Again, this matches other researcher’s findings for mastery learning and specifications grading in computer science courses (Bems, 2020; Morais et al., 2014; Sanft et al., 2021). This type of flexible assessment works best for upper-level courses that are not prerequisites for other courses because a student can pass without mastering all of the skills; their grade reflects the number of skills mastered rather than an average of the competency over all the skills taught. Other drawbacks of flexible assessment include, limited in-class preparation and review of programming assignments and an increase in the time required for grading. Providing the students with a required API for each programming assignment will help improve the grading times. Students have noted that they prefer specific guidance about when projects should be completed to prevent procrastination.

Acknowledgments

I would like to thank Kevin Peterson of the University of Lynchburg and the JMSCE reviewers for their valuable feedback on early drafts of this article. I would also like to thank my students and the faculty in the Computer Science Department at the University of Lynchburg.
References


Appendix A

Programming Assignment: Simple Key-Value Datastore Server Feature Set (11 pts)

Description
A key-value store is a simple database. It can store key-value pairs. A user can ask a key-value store to save a key like “ITR” associated with a value like “itrhelp@lynchburg.edu”. Later the user can ask the key-value store to retrieve the value for the key (ITR) to get back the associated value (itrhelp@lynchburg.edu). If you are familiar with maps or dictionaries, you can think of a key-value store as a program that behaves like a map or dictionary. You will be building a simple key-value store with a network API.

Grading
This project will be graded interactively in class or during office hours. When you are ready for me to grade it, let me know during class or sign up for office hours.

Required Features
- Server must accept serial connections, without restarting
- Server on cake/pie, client on desktop/laptop
- A client that demonstrates the features
- Client must have the following methods:
  - get(String key)
  - set(String key, String value)
  - put(String key, String value)
- You must be able to answer the following questions:
  - What is your message format?
  - What is your protocol?
  - How do you handle message framing?
    - How can you tell when you’ve received the entire message?

Features for Points
- 3 pts: set command
  - set a key-value pair
    - a get should be able to retrieve the value later
  - server should reject “” and null keys
    - server should tell the client the error
    - client should print the error message
  - server should reject keys that are already in the store
    - server should tell the client the error
    - client should print the error message
    - different error message from “” and null keys

- 2 pts: get command
  - given a key, provide the value
  - server should return an error if key is not in kv-store
    - server should tell the client the error
- client should print the error message

- 2 pts: put command
  - provide a new value for a key already in the kv-store
    - replaces the old value
    - a get should be able to retrieve the new value later
  - server should return an error if key is not in kv-store
    - server should tell the client the error
    - client should print the error message

- 3 pts: All UTF-8 characters are allowed in keys and values
  - including control characters like \n and \r

- 1 pt: Submitted code to Moodle as a zip file

**Not Required Features**

- Persistent connections
- Key-value store data does not need to be persistent
  - if the server is terminated, the data is lost
- Concurrent connections
- Keys or values that are not UTF-8 strings
- A nice user interface for the client
  - your client can just be code with no user input
  - it should output enough information so that I can be sure your server works

**Appendix B**

Rubric Example: Simple Key-Value Datastore Server Rubric Feature Set (11 pts)

**Prep notes**

- None

**Required Features (Double check these before starting)**

- Server must accept serial connections, without restarting
- Server on cake/pie, client on desktop/laptop
- A client to that demonstrates the features

**Questions to ask before starting**

- What is your message format?
- What is your protocol?
- How do you handle message framing?
  - How can you tell when you’ve received the entire message?
  - How can you tell when one message ends and another begins?
Features for Points

- **2 pts: simple set-get**
  - set “dog”->small
  - set “mouse”->alive
  - get “dog”
  - get “mouse”

- **2 pts: invalid sets**
  - set “”->small & set null->small
  - set “dog”->large
    - different message than “” and null keys
  - ask to see code where server handles invalid keys
    - no points if server is not the one that handles invalid keys

- **1 pt: invalid get: key not in the kv**
  - get “cat”
    - how can client differentiate between error messages and non-error messages
      - what if my value is an error message?
  - ask to see code where server handles invalid keys
    - no points if server is not the one that handles invalid keys

- **1 pt: put and get**
  - put “mouse”->”dead”
  - get “mouse”

- **1 pt: invalid put: key not in kv**
  - put “cat”->”small”
  - ask to see code where server handles invalid keys
    - no points if server is not the one that handles invalid keys

- **3 pts: All UTF-8 characters are allowed in keys and values**
  - set-get: "split\r\nkey" -> "split\r\nvalue"
  - set-get: "double\r\n\nsplit" -> "double\r\n\nsplit"
  - set-get: "period\r\n\n\nsplit" -> "period\r\n\n\nsplit"
  - ask how messages encoded
    - if using Strings, ask them how they delimited between cmd, key, value
    - what happens if that delimiter appears in key or value
    - if it cannot appear in key or value, no points

- **1 pt: Code in Moodle**
COLLABORATIVE LEADERSHIP FOR RESEARCH INVESTIGATING STEM TEACHER PREPARATION ACROSS MANY INSTITUTIONS

Dana Franz
Mississippi State University
dfranz@oire.msstate.edu

Keith Hubbard
Stephen F. Austin State University
hubbardke@sfasu.edu

Devon Brenner
Mississippi State University
dgb19@msstate.edu

ABSTRACT
This paper describes the creation of a collaborative research team investigating the impacts of education preparation on the recruitment and retention of science and mathematics teacher candidates in rural settings. Our collaborative research includes a core leadership team across 3 institutions and with collaboration across 14 total universities. We discuss the process from the inception through year two of this program, including the structure of leadership, communication techniques with the large group, and efforts to translate this research into scalable action. Using a framework for transdisciplinary research (Hall et al., 2012), we describe the processes and challenges that we encountered while engaging 14 institutions in a collaborative research project.

KEYWORDS
collaborative research, STEM teacher education, rural education

https://doi.org/10.25891/xzgf-td75
Over the past few decades, large-scale collaboration has emerged as a useful and informative model for research projects, particularly within the field of education. At the same time, as research requires input from and collaboration between a widely dispersed group of stakeholders, collaborative technologies have advanced in ways that enable groups across the United States and even globally to engage together in real-time. Examples of national efforts to foster this type of collaboration include National Science Foundation (NSF) Robert Noyce Teacher Scholarship (Noyce) Track 4 Collaborative Research grants (NSF, 2023a), the Association of Middle Level Education Research Initiatives (Mertens et al., 2015), and the recent call by the NSF S-STEM scholarship program to create research hubs that link multiple institutions across multiple sectors to “investigate evolving barriers to the success [of low-income students]” (NSF, 2023b). This paper examines the conception, grant proposal development, and initial enactment of a project supported by an NSF Noyce Track 4 Collaborative Research Grant entitled Teacher Preparation for Rural STEM Teacher Persistence and Retention or (TPR)². This 14-institution collaboration is focused on understanding the recruitment, preparation, and support of rural STEM teacher candidates as they complete educator preparation and become novice teachers in STEM classrooms; an endeavor that lent itself to a widely dispersed research collaboration involving multiple rural-serving educator preparation programs.

(TPR)² brings together multiple institutions, including large R1 universities and smaller teaching-focused schools, public and private institutions, as well as regional and main campus institutions, to jointly investigate a shared research question. Our multifaceted research design attempts to examine the impact of educator preparation programs by (a) understanding the ways that science and mathematics teacher preparation programs are addressing place and rurality, (b) exploring the connections between science and mathematics teacher education practices and beginning teachers’ willingness to seek out and accept rural teaching placements, and (c) studying how networking across institutions may provide insights into the recruitment and retention of science and mathematics teacher candidates in rural settings. A multi-institution collaboration was necessary for this project, not simply because the funding agency called for proposals for collaborative projects, but because of the interconnected nature of the work and the partnering institutions. Many of the institutions that serve rural STEM teachers are small in size or have a teaching-focused mission. These types of schools benefit from collaborating with institutions that have administrative capabilities that are adequate to undertake a research project of this magnitude. Also, smaller institutions have lower enrollment in their educator preparation programs and most likely would not, on their own, have enough data to lead to generalizable findings about educator preparation. It should be noted that even the largest programs in our project have only a handful of students enrolled in the various STEM education degree programs. Collaboration was also necessary because rurality by its nature means the programs being investigated were widely scattered geographically.

The obvious question, given the realities of rural research, is how might a team of researchers, STEM faculty, and teacher educators build the structure to make this collaboration work? This paper describes the creation of a research team that includes a core leadership team representing three institutions and collaborative partners across a total of 14 universities. We discuss the process from inception through year two of the program, including the leadership structure, the communication techniques we used for working with a large group, and our efforts to translate this research into scalable action. Using a framework designed by Hall et al. (2012),
we describe the processes and challenges we encountered while engaging 14 institutions in a collaborative project.

**Background**

Close to one-half of all U.S. school districts are rural or have schools that are classified as rural (Showalter et al., 2019). Research into rural education challenges is not new. Attracting and retaining STEM teachers to rural schools has been a pervasive problem for the past century (Biddle & Azano, 2016). As early as 1944, researchers were noting the marked differences between rural and non-rural schools (Dawson & Hubbard, 1944). Although there has been increasing attention to the crisis of teacher shortages in rural schools in recent years (e.g., Ingersoll & May, 2011; McHenry-Sorber & Campbell, 2019; Tran et al., 2018) staffing STEM classrooms continues to be a challenge for many rural schools across the U.S. Compared to urban and suburban schools, rural schools may offer lower salaries and have fewer resources for teachers. Other challenges to rural teacher recruitment include small communities facing their own economic challenges and higher poverty rates (Aragon, 2016). Perceptions of rural places may also contribute to recruitment challenges, including the persistent stereotype that rural students are somehow less capable and the widespread message that rural teaching is less desirable. While there are advantages to rural teaching that can include smaller class sizes, professional autonomy, community connection, and opportunities for leadership (Barley & Brigham, 2008), these advantages are not well-communicated, perpetuating the lack of teachers in rural settings (Goodpaster et al., 2012; Sutcher et al., 2016).

Similarly, there is a wealth of research on the increasing difficulties of recruiting, training, and supporting STEM teachers to any classroom, urban or rural. Colleges and universities continue to face decreasing enrollment in schools of education despite the increasingly sophisticated recruitment practices of many institutions. This is even though researchers with Get the Facts Out (2023) note that surveys of students in STEM programs indicate nearly half of all STEM majors have an interest in teaching. Institutions are forced to find non-traditional ways to support the dwindling numbers entering the field of education (King & Yin, 2022). For the past 20 years, NSF Noyce has provided funding to institutions of higher education to offer scholarships for STEM and STEM education majors who commit to teaching in high needs schools. The scholarship is an enhancement for students, but there are still barriers to recruiting a sufficient number of STEM teachers for the large number of high needs schools across the U.S.

Teacher shortages are also exacerbated by high rates of teacher turnover. In 2015, the National Center for Educational Statistics report (Gray & Taie, 2015) stated that nearly one-quarter of novice teachers leave their schools after three years. Teacher education accreditation standards have called for increases in content knowledge and higher entrance requirements with the hope of sending better prepared teachers into the field (Association of Mathematics Teacher Educators, 2017) and with the belief that increased entrance requirements might lead to more stability in the STEM teacher workforce. Teacher attrition rates are highest in rural areas, particularly in southern states and in schools that serve low-income and minority students (Anthony et al., 2017). Because teachers who persist in teaching become more effective over time (Papay & Kraft, 2015), attrition has a profound impact on student achievement (Darling-Hammond et al., 2017). Addressing persistence and retention in rural schools is a crucial step in providing equitable access to STEM teaching and learning for rural students (Behrstock-Sherratt, 2016; Eppley, 2016).
It may be that educator preparation programs can better prepare teachers for placements in rural schools so that they are more likely to consider a rural placement and, once hired, to remain in a rural school. We aimed to investigate how educator preparation programs address place and also the potential impact of specific features of educator preparation such as field placements, required readings, rural-focused scholarships, and other features, on new teachers’ initial placement and persistence in rural schools. Consistent with the NSF Noyce Track 4 call for collaborative research projects, (TPR)² aimed to create a research hub to collaboratively investigate these questions. This unique collaboration brought together both education and STEM faculty from 14 institutions to pool resources and share information about their programs in a truly multi-sector, trans-institution collaboration.

Collaboration across multiple institutions is complicated. Organizational and managerial research indicates the complexity of research when all partners do not hold the same titles, responsibilities, or credentials (Kieser & Leiner, 2011). The Carnegie Classification (Indiana University Center for Postsecondary Research, n.d.) provides a framework for designating types of institutions based on levels of degrees awarded, research activity in doctoral degree granting universities, size and type of enrollment, as well as specialized rankings such as tribal colleges. Large doctoral granting universities, often classified as R1 institutions, prioritize research while smaller regional colleges, or those that offer only undergraduate or masters degrees, prioritize teaching. Furthermore, multi-institutional research introduces the complications of each institution’s design, procedures for conducting research with human subjects, expectations for large-scale, grant-incentivized research, and faculty workload allocated to research endeavors. Thus, navigating the landscape of multiple institutions adds layers of complexity beyond the identified research agenda. Corley et al. (2006) discuss the need for either a highly organized research agenda or a well-developed organized collaboration structure for multi-institutional research to be successful.

Clearly, collaborative research is never easy. With this paper, we want to share the significant details, structures and supports that have helped us build a successful research team and contribute to a greater understanding of the features of multi-institution collaborative research that may support other projects.

Framework

Our research project is a collaboration between a wide variety of institutions from R1 universities to small regional colleges. Additionally, we have faculty that reside in either Colleges of Education or STEM departments in Colleges of Arts and Sciences. To understand the complexity of our work, we needed to draw upon the literature of teams. Hall and colleagues (2012) outline a model, drawn from team science, for conceptualizing a transdisciplinary framework in four phases: Development, Conceptualization, Implementation, and Translation. The Development phase of a transdisciplinary project entails convening a group of collaborators to examine a specific problem or area of interest. A Conceptualization phase is when the team collaborates to formulate hypotheses, conceptual models, research designs, or research questions to address the area of interest. During the Implementation phase the conceptualized plan is executed. Finally, a Translation phase consists of moving findings “from one level of analysis to another or across the discovery–development–delivery continuum in order to create innovative strategies for resolving or ameliorating societal problems” (Hall et al., 2012, p. 416).
This framework is explicitly articulated for \textit{transdisciplinary research}, which the University of Harvard’s School of Public Health, adapted by Young (2023), define as “research efforts conducted by investigators from different disciplines working jointly to create new conceptual, theoretical, methodological, and translational innovations that integrate and move beyond discipline-specific approaches to address a common problem.” They draw a distinction between transdisciplinary and \textit{interdisciplinary research}, which they define as “any study or group of studies undertaken by scholars from two or more distinct scientific disciplines. The research is based upon a conceptual model that links or integrates theoretical frameworks from those disciplines” (Global Arts and Humanities, Ohio State University, 2023) (see Figure 2).

Choi and Pak (2006) state that “[i]nterdisciplinarity analyzes, synthesizes, and harmonizes links between disciplines into a coordinated and coherent whole. Transdisciplinarity integrates the natural, social and health sciences in a humanities context, and transcends their traditional boundaries” (p. 359). Time will tell, but our research aspiration is indeed to transcend the traditional boundaries between rural education, STEM preparation, and classical educator preparation programming in a manner that addresses the preparation of rural STEM educators as a cohesive endeavor rather than several fractured parts. Additionally, Hall and colleagues (2012) argue that transdisciplinary research is distinctive in that it seeks to address societal problems and hence contains more traditional researchers as well as team members specifically included for their “relevant expertise to translate research findings into practice and policy applications” (p. 416). This is entirely consistent with our team design, which values research expertise in a variety of areas, but simultaneously includes team members particularly for their potential to translate findings and products into practical implementation, including faculty who work at teaching-focused and smaller rural-serving institutions. Hall et al. (2012) are explicit that these phases are typically not linear but cyclical, as our collaborative endeavors illustrate. Our research process will be interpreted through this framework.

\textbf{The Development and Conceptualization Cycle}

NSF Noyce aims to “address the critical need for recruiting, preparing, and retaining highly effective elementary and secondary mathematics and science teachers...” (NSF, 2023a). Tracks 1 and 2 of the program provides scholarships to individuals who seek to become STEM teachers—either through traditional undergraduate programs or through post-baccalaureate teacher preparation programs. Institutions design programs intended to recruit and support new STEM teachers and apply for funding. Track 3 grants provide funding to build leadership capacity and provide mentoring for current STEM teachers. Track 4 is intended to increase knowledge generation and translation by supporting collaborative research projects that investigate the programs and innovations being implemented in Noyce teacher preparation projects. Track 4 grants specifically call for teams to come together around a theme to investigate a shared research topic. It should be noted that Track 4 grants provide extra funding if a program has or previously had Track 1, 2, or 3 funding. This section describes how our team initially coalesced around the idea of submitting a Track 4 grant proposal.

\textbf{Gauging Interest in Rural-Focused Collaborative Research}

An early developmental step was realizing that rural teacher education might be a topic worth studying and gauging the interest of others in doing collaborative research. In the fall of
2019, two faculty at Stephen F. Austin State University (SFA)—a math professor and an education professor—who led Noyce programs for preparing teachers for rural Texas classrooms discussed how rural-serving Noyce teacher preparation programs were common, but how each individual program seemed to lack the scale to generate evidence about STEM teacher preparation that could move substantially beyond case study research. In order to gauge wider interest in rural STEM teacher preparation research, they developed a brief questionnaire about research interests and institutional connections to rural STEM teaching. Then, using the NSF Noyce project locator, they identified and contacted over 90 institutions that had programs that served rural students. From the perspective of transdisciplinary research, this was the initial development and conceptualization cycle in our work – two researchers believed that it might be important to collaborate to understand more about preparing rural STEM teachers, then used that work to identify a broader research team.

Of the institutions that were contacted, about 34 individuals responded with interest in collaborating on data gathering and research. Dozens of emails, phone conversations, and conference chats over the next two months led the SFA researchers to two conclusions: (1) that there was broad interest and willingness to examine the rural STEM teacher pipeline; and (2) as a regional, teaching-focused institution with limited research support for large collaborative projects, SFA was not an ideal institution to lead a group consisting of dozens of researchers, dispersed across the country, at least not at the level that this project deserved.

**Identifying a Core Leadership Team**

In their work to establish a broad team, the SFA duo also reached out to the National Rural Education Association (NREA) and began discussing the project with Devon Brenner, the editor of the organization’s journal and a rural education researcher at Mississippi State University (MSU). After several discussions with MSU, our team agreed that MSU would be well positioned to lead such a study but that it also would be wise to include Texas A&M University (TAMU) because of their researchers’ experience with Noyce programs and their methodological expertise in survey design and analysis—a key component of the project. The addition of research team members with expertise in distinct domains relevant to the key area of interest is one of the recommendations made by Hall and colleagues (2012) as part of the Development phase of transdisciplinary research projects. An important theme in cultivating these research connections is the significance of research associations. SFA researchers were able to connect with MSU researchers because of NREA, and MSU researchers had existing connections with TAMU researchers because of the Mathematics Teacher Educator Partnership – a network of mathematics teacher preparation programs begun in 2012 by the Association of Public Land Grant Universities.

This three-institution core leadership team began meeting regularly, formulating more precise goals and strategies, identifying previous research and a framework and logic model for a cross-institution research project, and building a shared understanding of the project’s research goals and the NSF Noyce Track 4 funding mechanism. In Table 1, we outline the collaborative strengths and weaknesses of our lead institutions and their research teams.
Table 1
Collaborative Strengths and Weaknesses of Lead Institutions and Research Teams

<table>
<thead>
<tr>
<th>Institution</th>
<th>Collaborative Strengths</th>
<th>Collaborative Weaknesses</th>
</tr>
</thead>
</table>
| MS State    | • Expertise and connections in rural education, including rural teacher shortages  
             • PI connected to NREA and editor of The Rural Educator and author of multiple seminal pieces about rural teacher preparation  
             • Co-PI had experience with Noyce Track 1 and Capacity Building projects and other NSF teacher education grants  
             • MSU Classified as an R1 institution with institutional resources to support collaborative research  
             • Office of Research Compliance manages multiple multi-institutional projects needing IRB approval | • PI is less experienced in STEM research  
             • Leadership has many other administrative duties |
| SFA         | • Experience with Noyce Track 1, 3, and 4 projects  
             • Experience with studying rural STEM teachers  
             • Network of dozens of rural STEM teachers | • Lacks large-scale research support structure  
             • Lacks experience coordinating multiple institutions |
| TAMU        | • Experience with Noyce Track 1 and 4 projects  
             • Expertise in longitudinal STEM teacher education research and publishing  
             • Classified as an R1 institution with institutional resources to support collaborative research  
             • Evaluator with expertise in STEM research | • Leadership has many other administrative duties  
             • IRB approval system is complex |

Developing a Focus for Research Collaboration

The research coalesced around two central data collection components: (1) examining the Educator Preparation Programs (EPPs) at universities involved in the study, then (2) surveying prospective teachers who had participated in these EPPs over a sustained period to understand whether they had taken jobs as teachers, whether their teaching jobs were in rural schools, and whether they persisted in remaining in the teaching field and persisted in teaching in rural classrooms. After discussions with potential partners, over time we settled on proposing the following goals:
Goal 1: To investigate the impact of EPP programmatic features on program completers’ intention to teach as well as their persistence (continuing to teach at the same school) and retention (continuing to teach but at a new school) in rural STEM classrooms.

Goal 2: To engage in deep reflection leading to programmatic adjustments within collaborating partners’ EPPs intended to increase equitable access to effective instruction for rural schools.

Goal 3: To share emerging knowledge about programmatic features of EPPs that support program completers’ intention to teach, their employment decisions after completing the EPP, and their persistence and retention in rural placements.

Identifying Collaborative Partners

Another theme within transdisciplinary research is identifying not just the research challenges but the logistic and human challenges to long-term collaboration, then forming a team and a strategy to address those challenges. Accordingly, as the core research team developed a clear understanding of the types of engagement that would be needed from partner universities, the vision was communicated to the 34 institutions who had expressed initial interest in collaborating on a rural teacher education research project. A key task was to build the full research team, including determining which institutions not only had an interest in the research (and in securing the funding) but also would have the ability and the staying power to work through their local institutional review board (IRB) and other research requirements, participate fully in the proposed project, acquire the data necessary for analysis, and continue the research for the multiple years required to conduct a longitudinal study such as this one.

For the core leadership team, it was important for the whole team to keep the long view in mind – having “partners” who would not be able to follow through on all of the commitments would detract from the entire project and make success far less likely. Frankly, we aimed to identify collaborating faculty and institutions that would readily reply to communication (email) and provide data to support the collaboration. One strategy we used to identify partners was to request information that would be helpful for proposal development. We asked potential partners to report the numbers and demographics of students who had graduated from each of their programs over the past year—this information was necessary for the development of the proposal, but asking for data about their programs was also an opportunity for institutions to demonstrate their institutional capacity to identify and share program data in a timely fashion. As the project's details were developed, a few institutions determined they were unable to complete the project and chose not to participate. Eventually our team consisted of 11 partner universities in addition to the three core universities—the maximum number of collaborating partners allowed by the Noyce Track 4 funding mechanism and the number of institutions expressing interest in developing the joint proposal and participating in the 4-year project. See Table 2 for a complete list of partners and their roles.

Seeking Funding

The culmination of the conceptualization phase of the project was formulating a Noyce Track 4 collaborative grant proposal. We began the work of writing research questions based on
Table 2
Partner Institutions and Collaborators, Along with Their Roles

<table>
<thead>
<tr>
<th>Institution</th>
<th>Affiliations of Collaborators</th>
</tr>
</thead>
</table>
| Alabama A&M University                   | Department of Physics, Chemistry & Mathematics  
|                                          | Office of Teacher Education and Certification                                                 |
| Clarkson University                      | Department of Mathematics  
|                                          | Department of Education  
|                                          | Department of Physics  
|                                          | Institute for STEM Education  
|                                          | School of Engineering                                                                       |
| Fort Hays State University               | College of Education  
|                                          | Department of Physics  
|                                          | Science & Education Institute                                                               |
| Mississippi State University             | Department of Biological Sciences  
|                                          | Department of Education  
|                                          | Office of Academic Quality  
|                                          | Social Science Research Center                                                               |
| Morehead State University                | Department of Middle Grades and Secondary Education  
|                                          | MSUTeach                                                                                     |
| North Dakota State University            | Department of Biological Sciences  
|                                          | Office of Teaching and Learning  
|                                          | School of Education                                                                         |
| Stephen F. Austin State University       | Department of Mathematics & Statistics  
|                                          | Department of Education Studies                                                             |
| Texas A&M University, Commerce           | Department of Physics and Astronomy  
|                                          | Department of Curriculum and Instruction                                                    |
| Texas A&M University, College Station    | AggieTeach – Arts & Sciences  
|                                          | College of Arts & Sciences  
|                                          | Department of Mathematics  
|                                          | Department of Teaching, Learning, and Culture  
|                                          | Education Research Center                                                                   |
| Texas Tech University                    | Center for Integration of STEM Education & Research  
|                                          | Center for Transformative Undergraduate Experiences  
|                                          | College of Education                                                                        |
| University of Alabama at Birmingham      | Center for Community Outreach Development  
|                                          | Department of Biology  
|                                          | UABTeach                                                                                     |
| University of Kentucky                   | Department of STEM Education                                                                |
| University of Wisconsin River Falls      | Center for Excellence in Teaching and Learning  
|                                          | Department of Physics  
|                                          | Department of Continuing Education                                                           |
| Winthrop University                      | College of Education, Sport, and Human Sciences  
|                                          | Department of Biology  
|                                          | Department of Mathematics                                                                   |
the funding agency’s Request for Proposals and the defined scope of work. Each of the co-PIs brought a piece of the puzzle to the collaborative work—experience with the rural teacher preparation and the Noyce scholarship program, quantitative or qualitative research expertise, teacher preparation experience, and project management experience. It was important that we have explicit discussions about who would take responsibility for each task. We also identified an evaluator and included her in our discussions early in the process.

**Facilitating Communication**

The media tool used for collaboration, Zoom, and the way it was utilized, was also fundamental to the project. A platform for connecting remotely was absolutely necessary given the disperse geography of the 14 institutions. Zoom allowed for a certain degree of community building and added visual, voice, and document sharing facets to our collaborative efforts. The team was very intentional about the structure of Zoom meetings, particularly with the full group of 14 institutions, discussing both content and relational aspects of the agendas, engaging participants in smaller break-out groups, etc. It is likely that the circumstances that necessitated the decrease in in-person meetings during the COVID-19 pandemic strengthened our collaborative efforts since the intense time of formulating the collaboration was in the spring and summer of 2020, when most states were locked down.

Online tools like Zoom and the implementation of COVID-19 protocols put collaboration with colleagues across the country on an equal footing with collaborating with colleagues who work down the hall. For our project, all core leadership meetings take place through Zoom and have a clear agenda that is set prior to meeting. We designate a specific hour each week during each semester for our meetings and cancel meetings only for holidays or if most of the core leadership members are attending a conference. Each Monday, the program manager (described in more detail below) provides a recap of the previous week’s project-based work as well as information for consideration or items to complete before the next meeting. Whole group meetings (i.e., meetings that include all 14 institutions) are held once a month. Similar to the leadership meetings, an agenda and updates are sent to the whole team every few weeks.

Equally important is the document and information sharing within collaboration. Since communication is central to all four phases of transdisciplinary research, choosing a method and a medium for synchronous communication was vital. The core leadership team utilizes Microsoft Teams as the primary document storage platform. We found additional platforms just added confusion. Our program manager assists all members with storing and locating documents and data.

We close this section on the conceptualization and development process that we followed with a brief vignette highlighting how a transdisciplinary team with diverse stakeholder connections enriched our research.

**Vignette – Expertise and Learning from Each Other**

Our original conceptualization of the engagement of prospective STEM teachers in an investigation of place centered on Azano and colleague’s (2019) 3 Cs model for preparing rural teachers – focusing on Curriculum, Context, and Conveyance. The 3 Cs framework was co-authored by the PI. This model makes explicit both the importance of bringing to the forefront issues of preparing teachers for potentially teaching in rural places, and it seeks to intentionally
identify the different domains that contribute to this work. Designed by rural teacher education researchers, the framework does not yet appear to have permeated thinking in other disciplines – specifically for our purposes, the field of STEM educator preparation.

The 3 Cs framework emphasizes the need for placed-based teacher education experiences to develop a preservice teacher’s understanding of the importance of the interplay of context, curriculum, and conveyances in a rural education classroom. An example of context in small rural schools is when teachers are assigned non-traditional tasks or asked to fulfill multiple roles including teaching multiple grades, teaching a variety of courses within or across content areas (e.g., teaching calculus and physics), or sponsoring extra-curricular activities potentially even during their first year of teaching. Programmatic features related to curriculum include instructional practices that focus on place and leverage of local resources for mathematics teaching (Avery, 2013). Programmatic features related to conveyance focus on access to teacher education programs. Conveyance includes recruitment strategies (e.g., focus on rural residents, partnerships with rural high schools, or grow-your-own programs) and flexible offerings for completing the educational requirements for licensure.

After funding, first our core leadership team, then our full 14-institution transdisciplinary team began to grapple with the 3 Cs framework in greater detail. When the mathematicians and scientists began to engage with the definitions of Curriculum, Context, and Conveyance as they are presented in Azano et al. (2019), they were much more concerned with precisely delineating the categories than the original authors had been. This led to a fruitful discussion, which persisted over several months about what really constituted each area of the framework, how they differed, and even whether the category definitions should be refined slightly to fully communicate the meaning to stakeholders across disciplines. Engaging in this type of collaboration took tremendous humility on the part of the rural education researcher on our team. In particular, a willingness to allow researchers from outside of her discipline to challenge established frameworks within her discipline. But the result has been a thoroughly engaging exercise that has pushed forward both the understanding of place-focused teacher education and our efforts to code the data (e.g., lesson plans, syllabi, interview transcripts, etc.) collected for this study as belonging to one or more of the 3 Cs categories. Transdisciplinary teamwork has informed our conceptual framework so that a broader group can engage deeply with it, and we can use it to pursue our research goals more effectively.

**Implementation in Team-Based Research**

The implementation process interacts cyclically with conceptualization and development informing how the project is conceptualized and guiding development. Early in our work, the team engaged only in the three areas depicted in Figure 1 (see Hall et al., 2012, p. 417). The development and conceptualization phases were iterative and implementation was our initial goal. Implementation is the phase of a project where a shared understanding transitions into fully developed functional roles – who knows what, who does what, how things get done, and how interactions occur (Mesmer-Magnus & DeChurch, 2009).

**Team Communication**

The core leadership team determined that regular communication was absolutely essential in keeping our cross-disciplinary team collaborating. As soon as NSF funded our project, the
leadership team resumed the weekly Zoom meetings that had been the hallmark of the development and conceptualization phases of the project. It should be noted that the core leadership team has been very consistent with the continuation of weekly meetings which we believe has helped us to maintain our research-focused momentum.

Whole team meetings and regular email updates provided both the onboarding that was necessary at the beginning of the project and essential ongoing support for collaboration over the first two years of the project. Whole team meetings involving at least one PI or co-PI from each of the 14 partner institutions have been held 2-3 times per semester (about every other month). While monthly whole team meetings were originally planned, due to the ebbs and flows of developing surveys and collecting data it became evident that monthly meetings were not necessary and potentially could be seen by attendees as just one more obligation. We worried that unless there was a real need to meet—to get feedback on survey development, or to share plans for data collection, for example—meeting for the sake of meeting might have the unintended consequence of diminishing participation.

In addition to virtual meetings, we meet as a whole team in person once each year. We “all but require” attendance at a half-day meeting prior to the annual PI meeting for all Noyce projects which is held in DC each summer. Travel is already built into the project budget, so collaborators should have minimal hindrances to attending. Our in-person gathering is used for community building and developing a shared understanding of the project framework, timeline, and goals for data collection, analysis, and dissemination.

We also needed effective avenues for asynchronous communication. We designed a webpage to share information both externally and within the team (with password protection). We needed team members to have ready access to information and collaborative components of the project and to be able to upload documents and data in a secure format. Most importantly, the platform needed to have an organizational structure that people could understand and easily access. Unfortunately, the website approach, though conceptually valid, did not end up implementing well as some team members were unable to access secure files, and others had trouble understanding where different files were located. We attempted other collaborative file-
sharing solutions before settling on OneDrive and Microsoft Teams since these platforms are supported by many of the universities involved and are more secure than other formats we investigated. This is another small example of the iterative nature of collaborative research. Ultimately, data management and the handling of diverse research artifacts was best facilitated by having an experienced researcher assume the role of Program Manager (PM) along with seasoned qualitative and quantitative researchers on the team serving as co-program managers.

**Team Facilitation**

The PM is a vital member of our research team. Finding someone to serve in this role to complement the research strengths of other team members was essential. A researcher with a PhD in biology, an interest in STEM education, and experience coordinating large teams serves as PM and strengthened the team in two specific ways. Although the core leadership team members had expertise in rural education, educator preparation, mathematics and mathematics education, program evaluation, and both qualitative and quantitative research methods, the team lacked an individual with experience in the natural sciences. The PM added this important research perspective, and, perhaps more importantly, brought translational benefits to the project collaborative experience in this area and the ability to translate our research findings in ways that are significant to natural science faculty across the 14 institutions as well as to a broader audience.

To date, the PM has collected data about educator preparation programs, including interviews with faculty and administrators and documents such as course syllabi. She has cataloged and organized the data in such a way that the leadership team can access information at any time. She has also distributed surveys to teacher candidates and program completers and compiled survey responses. She is currently leading the beginning steps of data analysis. The PM’s job, however, is not just to do the research collection and analysis, but rather to facilitate the team as a whole in the process. This is an essential step in implementation – allowing the research work to truly be distributed across the team. The PM collaborates with the core leadership team as well as with team members and stakeholders at all 14 institutions. For example, once hired, the PM immediately began systematizing the communication process and data storage procedures. Summary emails are sent by the PM to members of the core leadership team every Monday, recapping the previous week’s work, setting the week’s agenda as well as providing materials that should be considered by the team prior to the Thursday Zoom meetings. The PM keeps an eye on our overall timeline and scope of work to ensure that project deadlines are being met. She schedules, plans, and sends out summaries of whole team meetings, and sends regular email updates about the current status of data collection and analysis to keep the entire team informed about the project.

**Overcoming Obstacles**

Distributed research involves challenges that are distinct from more traditional research. One example is the process of obtaining IRB approval for all 14 institutions to participate in data collection for the (TPR)² project. Our research plan emphasized having the PM and the research team at MSU conduct the majority of the data collection including interviews and survey distribution. Therefore, we planned to use a single IRB submission for collaborative research through the MSU Office of Research Compliance. Although we were conducting research
methods that are standard educational practice and ultimately the project was determined to fall in the “exempt” classification for human subjects research (see Health and Human Services, n.d.), the process of obtaining IRB approval took months longer than we anticipated. Further, of the 14 institutions awarded NSF funds through the project, only 12 institutions were able to move forward with approval under the central MSU IRB application. Another two institutions spent months attaining IRB approval at their own institutions. One of those institutions concluded that only university employees may distribute the surveys and communicate with participants, while the other ruled that the IRB process that was in place was sufficient. We needed an understanding of institutional policies and procedures as well as persistence to overcome this obstacle. We were not anticipating the length of time nor the complexity of the task that was necessary to get approval for all institutional partners.

**Putting it All Together**

Once the team was in place and the collaboration and facilitation obstacles had been addressed, we were ready to move forward with the work of the project. The next section highlights how the logistical and organizational planning for implementation and conceptualization has resulted in meaningful collaboration across multiple institutions. In fact, we believe the work of our project is stronger because we are collaborating.

**Collaboration in Action**

We have selected one story to highlight how the diversity of our research team in a variety of ways including research specialties, academic roles, and geographical locations has influenced the project. Although the activities are squarely a part of the implementation phase, they also overlap significantly with conceptualization and development phases, and even the translation phase (discussed below) as described in Hall et al. (2012).

The development and implementation of surveys was central to the research goals for our project. We identified three distinct surveys that we planned to create and distribute to gain longitudinal insights into prospective rural STEM teachers plans and decisions at the end of their teacher preparation program. For this purpose, we developed the Teaching Intention Survey as well as the Initial Employment Survey, completed at the beginning of the first year and after program completion, and the Teacher Follow-Up Survey, completed a year after program completion.

The content and validation techniques of these instruments are described in more detail elsewhere (e.g., Whitfield et al., under review). Here we will focus on the role of these surveys in our project’s collaborative efforts. A first draft of each survey was designed by the lead research team and shared with all partners before we submitted the initial grant proposal to NSF. At the time our project was developed no validated survey existed in the literature for exploring teachers’ views about teaching in rural, and, since they were in draft form, our survey items had not been evaluated for face or content validity. Once the grant was funded it became clear that the draft surveys needed a great deal of revising in order to align with our research questions and provide valid data about teachers’ intentions and views of rural teaching and teacher preparation.

The first noteworthy obstacle was that we aimed to determine whether graduates of our programs sought out or accepted rural teaching positions and to understand their views about teaching in rural schools. However, we found that we had many questions about what counts as
rural. Across our team, we met several times to discuss “rural”. Whether, for example, a small college town was “as rural” as a farming community or whether students from forested areas of upstate New York or the plains of Kansas would consider themselves “rural.” Much of our initial work to revise and finalize our surveys focused on discussions about defining this word.

The federal government has at least 15 different definitions of rural in various agencies and funding programs (see Washington Post, 2013) and rural researchers themselves do not always use a concrete definition of the word rural (Longhurst, 2022). We also debated whether to classify places using terms like urban, suburban, rural, or remote, to distinguish between distance from urban areas or population density. This was familiar, and comfortable, territory for the educational researchers on our team, but not for the biologists and mathematicians who expected a single, concrete definition of rurality. After repeated discussions, listening to team members and their thoughts about rurality and the students and communities they serve, and reading some rural education scholarship on defining rural (e.g., Brenner, 2022; Longhurst, 2022) we settled on two terms—rural and not-rural—and a definition of rural for our surveys of “a community smaller than 50,000 people” with the hope that this definition would allow survey respondents to consistently answer survey questions about where they are from and where they are teaching. It should be noted that the definition we settled on is derived from the multiple definitions put forth by federal agency literature.

Similar in-depth conversations about terminology, survey item wording, and survey focus influenced the development of the final versions of other survey items as well, including questions related to course assignments and experiences in teacher preparation. Having dozens of researchers in different contexts provide multiple rounds of feedback allowed us to understand which questions might be interpreted differently by survey respondents. This iterative and collaborative review process also allowed us to evaluate the connections between different survey prompts and our project research questions from different vantage points. Since the target audience for completing the surveys was science and mathematics majors, it was important to cover significant educational ground but to do so in a way that would make sense to those with STEM training and research interests.

Most significantly, all 14 institutions had to be sufficiently aware of and committed to implementing the final versions of the surveys. With their understanding and support, we felt they would encourage students to take our surveys ensuring good response rates. Again, this was a collaborative effort. The surveys were customized by the PM. For most sites, IRB allowed student contact information to be provided to the PM so that the surveys (and reminders) could be distributed centrally. However, for two sites, the IRB panels determined that the surveys had to be distributed internally. Response rates on the second round of the Teacher Intention Survey were distinctly lower than the first. Luckily, one of the PM’s roles was to keep abreast of these response rates. The core leadership team reflected on the response rate issues and chose to develop a solution as a collaborative effort. Rather than treating certain sites as ‘problems’ that needed to get in line with a proposed solution, all sites gathered, discussed response rate issues, and explored solutions collaboratively. Different institutional contexts required different solutions. At one institution all of the potential respondents gathered for a particular class so the survey link and reminders could be shared during that class. At another institution, it was best to have a person other than the professor of the course distribute the survey. We believe collaboratively approaching the response rate concern not only helped generate a broader range of potential solutions for a variety of contexts, but also increased the engagement of institutions in the overall project.
Translation

The translation phase of collaborative research also interacts cyclically with the other phases, as depicted in Figure 2; although translation, as our team is designed, was the last phase to be included in the cycle. The figure depicts the entire interactive model. However, given the nature of our work and the time necessary to collect data across 14 institutions, after three years, our project is entering into the translation phase.

Figure 2
Four Phase Model of Transdisciplinary Research (Hall et al., 2012, p. 417)

Currently, the (TPR)² project is about half-way through its funding cycle, so we see ourselves as beginning to move to a more translational focus. Year One surveys have been collected and are in the process of being analyzed. Interviews of key stakeholders have been completed and transcribed and we are in the process of coding and interpreting the documents. The core leadership team as well as other partners have presented at several conferences on the need for rural STEM teacher education research, the structure of the collaboration as a promising practice, and preliminary findings based on our analysis student survey data to date. Through these efforts, we are working to communicate what we have initially learned about the significance of considering place, while also extending an invitation for rural-serving educators to join our collaborative efforts. Additionally, our research team has written two blog posts and is in the process of writing three collaborative papers on topics such as the early results of place-based STEM education methods, large-scale survey development, and the implications of influencers on the selection of schools or districts where graduates who are seeking teaching jobs are accepted. Partners are teaming with core leadership team members to investigate other topics of interest that have arisen during our whole team meetings. We are working to raise awareness of the significance of place in educator preparation, to increase the exchange of ideas about curricula and contexts that foster this type of place-based educator preparation, and to establish connections with university stakeholders who are interested in moving forward with this work.

While interest is evident, translation implies that the research results impact practice. It is too early in our project to measure this effect. However, the framework does seem to have implications for collaboration in educational settings.
Conclusion

At the midpoint of the funding of our project, we appear to have a successful collaboration that is leading to meaningful transdisciplinary research. The intentionality of building and enacting collaboration is key to continuing our project and eventually to our research findings having a transformative effect on the fields involved. The transdisciplinary framework developed by Hall and colleagues (2012) provides a structure for understanding why our work can be used as a model for building and understanding collaborative research. As we move further into the translation phase, there will be an emphasis on shared research, shared conference presentations, and other collaborative opportunities. The initial translation stages indicate that our group will be able to continue to work collaboratively in meaningful ways.

It was important for the establishment of our team to spend a significant amount of time in the development and conceptualization phases. As suggested by Corley et al. (2006), we needed to have a crystalized picture of what we hoped to accomplish. We were able to develop and implement a research project because we took the time to make sure each member of the core leadership team had a shared vision of the work. As described in this paper, we had to navigate the differences of our team members’ research interests, the positions we held at different universities, and the administrative and procedural complexities of our universities even though two of the three lead institutions are classified as R1 institutions. Only after the significant work of development and conceptualization could we begin the process of implementation. And, in our case, writing a grant that was selected for funding with a defined scope of work that indicated to reviewers at NSF that we would accomplish this work.

The cyclical and iterative processes of the Hall et al. (2012) framework helped us realize our collaboration is successful. Each time we experience challenges, we rely on the structure of our team to help us navigate to a solution. The key components for our team development appear to be:

- the intentionality of selecting members of the core leadership team,
- the significant amount of time spent on shaping the goals of the research work,
- the careful assembling of the larger research team with clear messaging about the work and participation expectations,
- well-defined self-reflection and self-correction plans as demonstrated by communication procedures and survey development,
- the selection of the program manager with specific organizational skills, and
- organized, regular meetings of both the core leadership team and the larger whole group.

We openly acknowledge that our core leadership team has developed into a friendship over time. This friendship has enhanced our work time as we share the successes and challenges of the project. While team building as well as opportunities to get to know each other were deliberately part of the development and conceptualization phases, the true friendships that have evolved were an unforeseen but welcome bonus that emerged from the process.

By sharing our process, our successes, and our challenges in this article, our intent is to allow other teams working at the intersection of STEM, education, and rurality to better approach the large collaborations necessary to affect the broad change called for by Biddle and Azano (2016) in addition to many others. Sustained, multi-institution dialog and the exchange of best practices seems to be the only reasonable path for improving rural STEM educator preparation and support. This transformative work is clearly necessary if our rural communities are to have equitable access to high quality STEM education.
Acknowledgment

This paper was developed as part of the Collaborative Research: Investigating STEM Teacher Preparation and Rural Teacher Persistence and Retention project and is supported by the National Science Foundation, lead awards 2050099, 2050108 and 2050249. The opinions expressed here are those solely of the authors and do not reflect the opinions of the funding agency.

References


Global Arts and Humanities, Ohio State University (2023, April 9). *Why cross-disciplinary research matters*. https://globalartsandhumanities.osu.edu/research/cross-disciplinary-research


Washington Post (2013, June 8). *The federal definition of rural—times 15*. [URL: https://www.washingtonpost.com/politics/the-federal-definition-of-rural--times-15/2013/06/08/a39e46a8-cd4a-11e2-ac03-178510c9cc0a_story.html]
