2024

Examining the Influence of Secondary Math and Science Teacher Preparation Programs on Graduates' Instructional Quality and Persistence in Teaching

Danielle M. Rhemer
Old Dominion University, drhemer@odu.edu

Will Rogers
Yale University

Sherry Ann Southerland
Florida State University

Original Publication Citation

Rhemer, D. M., Rogers, W., & Southerland, S. A. (2024). Examining the influence of secondary math and science teacher preparation programs on graduates' instructional quality and persistence in teaching. *Education Sciences*, *14*(5), 1-28, Article 506. https://doi.org/10.3390/educsci14050506 *education sciences*

Article

Examining the Influence of Secondary Math and Science Teacher Preparation Programs on Graduates' Instructional Quality and Persistence in Teaching

Danielle Marie Rhemer [1,*], Will Rogers [2] and Sherry Ann Southerland [3]

[1] Darden College of Education & Professional Studies, Old Dominion University, Norfolk, VA 23529, USA
[2] Department of Ecology and Evolution, Yale University, New Haven, CT 06511, USA; will.rogers@yale.edu
[3] School of Teacher Education, Florida State University, Tallahassee, FL 32306, USA; ssoutherland@fsu.edu
[*] Correspondence: drhemer@odu.edu

Abstract: This quantitative, non-experimental study explored the relationship between the features of math and science teachers' preparation programs and their graduates' instructional rigor and persistence in teaching. Five math and science teacher preparation programs from across the United States were examined. Six sets of instructional tasks were collected from forty-six recent graduates of these programs to provide insights into novices' instructional rigor, and employment data were collected for thirty-seven of these graduates three to eight years after graduation. Regardless of the program's features, all teachers could design and implement instruction with moderate to high rigor. However, this ability was not the norm. Mixed-effect models suggest the strongest evidence between degree types (bachelor versus post-bachelor) was related to teachers' persistence: novices from graduate programs were more likely to persist in the work. However, no program feature was strongly associated with instructional rigor. Further research is needed to determine if the differences we found in teacher persistence are due to the nature of applicants drawn to particular programs (undergraduate versus graduate) or the program's structure. Future research is also needed to explore the influence of instructional context (i.e., district, school, and department norms for instruction) on math and science teachers' instructional rigor.

Keywords: science; math; teacher preparation; teacher persistence; instructional rigor

check for updates

Citation: Rhemer, D.M.; Rogers, W.; Southerland, S.A. Examining the Influence of Secondary Math and Science Teacher Preparation Programs on Graduates' Instructional Quality and Persistence in Teaching. *Educ. Sci.* **2024**, *14*, 506. https://doi.org/10.3390/educsci14050506

Academic Editor: Daniel Muijs

Received: 3 March 2024
Revised: 19 April 2024
Accepted: 30 April 2024
Published: 8 May 2024

1. Introduction

Teachers are the most important in-school factor in improving student outcomes [1–3]. Quality teachers can raise the achievement levels of their students, while poor ones hinder the learning of those in their classrooms [4], and this relationship holds when one examines the impact of mathematics (math) and science teachers, specifically [5]. The current policy environment in the US makes skilled math and science teachers more critical than ever because of the increased rigor of new standards in science and math, together with a continued focus on teacher accountability for student learning [6]. Current instructional reforms, such as the Common Core State Standards in Mathematics [7] and the Next-Generation Science Standards (NGSS) [8], emphasize developing students' deep understanding of key ideas by engaging them in disciplinary practices, and the success of these reforms depends mainly on the quality of classroom instruction [9]. Undoubtedly, as Cohen and Ball [10] argue, "policies that seek to change instructional practice depend upon—and are changed by—the practice and the practitioners they seek to change" (p. 238). Therefore, teachers are the critical agents in putting such ambitious instructional reforms into practice, and to succeed, these reforms require shifts in the way science and math are currently taught in many classrooms [11,12].

However, worldwide, there are documented shortages of teachers [13,14], something that has accelerated in the wake of the COVID-19 pandemic. Oluk [14] (p. 1) calls our attention to shortages in teachers and trends in teacher retention rates as a "major cause[s] for concern". The "teacher shortage" has received a great deal of attention in the US, even before the pandemic. In the years after the pandemic, teacher vacancies have increased substantially, with some reports showing a 50% increase in teacher vacancies from 2021 to 2023. Additionally, 9 to 10 percent of positions in the United States are filled by teachers who do not have a license or teach outside their subject area [15,16].

Over the past three decades, there has been an emphasis on understanding the central causes of teacher shortages in the United States to reverse these trends [17,18]. The findings of this work highlight that teacher shortages are more pronounced for secondary math and science teachers than for teachers of other subject areas and are exacerbated in high-needs settings (e.g., worse working conditions, high historical attrition rates, etc.). The number of secondary math and science teachers leaving the profession for reasons other than retirement is high, and schools that serve disadvantaged populations of students experience the highest turnover [19,20]. This revolving door of math and science teachers, particularly in low-income and high-needs schools, can create teacher shortages and exacerbate achievement gaps. In response to these documented teacher shortages, there has been an expansion of how teachers are being prepared and certified to teach, from various alternative pathways to more traditional university-based, preservice teacher education programs [21]. Despite the proliferation of teacher preparation pathways, teacher shortages in secondary math and science continue and are forecasted to remain in the future [18].

The entire field of teacher education is posited on the notion that how preservice teachers are prepared for the work of teaching will influence their instructional approaches and their skills within those approaches [22–24]. Preservice teacher preparation routes vary, and two blunt distinctions include traditional versus alternative teacher preparation programs [25]. "Traditional" teacher preparation programs are generally offered through a college of education as part of a four-year undergraduate degree and include a mixture of disciplinary- and pedagogy-specific coursework, as well as work in classrooms under the guidance of mentors. "Alternative" programs can vary—from short summer programs that place candidates in classrooms after a few weeks of training to those that offer 1–2-year post-baccalaureate programs. Both alternative and traditional programs have entry requirements (99.5%), such as high school and/or college transcripts (96.5%), minimum GPA (95.3%), and a minimum number of courses/credits/semester hours completed (84.9%) [26]. Though the entry requirements are similar, recent data show that traditional teacher preparation programs are the most common nationwide (81%), with this trend continuing for math (77%) and science teachers (70%) [27]. Additionally, traditional teacher preparation programs have been found to produce teachers who are more qualified, confident about their preparedness to teach, and have higher entry and retention rates [22,28,29]. However, research findings are highly variable, and the influence of the characteristics of programs that prepare teachers on their graduates' persistence and instructional quality is poorly understood [30].

Considering this gap in understanding regarding what characteristics are necessary to train quality teachers who can persist in their work and the importance we place on all students obtaining a quality education, we have chosen to focus our study on traditional secondary math and science teacher preparation programs that support their graduates to work in high-needs settings; specifically, exploring the relationship between secondary math and science teachers' preparation, instructional quality, and persistence in the field. To explore this relationship, we examine the characteristics of five secondary math and science preservice teacher education programs across the United States. We analyze whether these characteristics influenced the rigor of their graduates' teaching and their persistence in the field. The study's findings have implications for traditional preservice teacher education programs, highlighting changes we can make in how we train secondary math and science teachers to help them persist in the field of education.

2. Related Literature

This article explores the relationship between secondary math and science teachers' preservice preparation, instructional quality, and persistence in the field. Thus, the following section will unpack what is known about these areas of study.

2.1. Research into Features of Teacher Preparation Programs

Teachers with little to no preservice preparation leave the profession at much higher rates than those with more preparation [31]. Beginning teachers praise preparation programs where they spend extensive time in schools and feel better prepared because of the more extensive pedagogical coursework [32–34]. Indeed, teachers' perceptions of the quality of their preparation are linked to plans to remain in the classroom [32,35]. However, close examinations of the structure of teacher preparation programs in terms of their influence on instructional quality and teacher retention have yielded sometimes contradictory results [34].

Research on teacher preparation and retention has primarily focused on differential retention rates across preparation routes or programs [36]. Depth of teacher preparation is one of four major factors contributing to teacher attrition [37]. However, little empirical work examines the relationship between teacher preparation features and retention. Using data from the Schools and Staffing Survey (SASS) and subsequent Teacher Follow-Up Survey (TFS), Ingersoll, Merrill, and May [38] examined the relationship between preservice pedagogical preparation and teacher attrition. They found that after their first year of teaching, math and science teachers with more coursework in teaching methods were less likely to depart, as were teachers whose coursework included learning theories and preparation in selecting instructional materials. However, the same relationship was not found for teacher effectiveness in a study by Harris and Sass [39], as they described content and pedagogy courses as mainly unrelated to a teacher's instructional quality (as measured by standardized assessments) and suggested that pedagogy coursework can sometimes be negatively correlated with instructional effectiveness.

A review of the literature on teacher preparation programs provided mixed support for the importance of subject matter knowledge; however, the extant literature base lacks sufficient detail to indicate how much subject matter coursework is most beneficial or at what point diminishing returns to achievement gains set in [40]. Internationally, secondary teachers take twice as many math courses in countries with the highest mathematical content knowledge [41]. In the US, there is evidence that math coursework is positively related to instructional quality [42,43], but other studies found that math courses are negatively related to instructional quality [44]. Preservice high school teachers who received more subject matter coursework have been shown to have more significant increases in content knowledge throughout their preparation programs [45]. Still, research has not addressed the relationship between subject matter preparation and retention.

In contrast to the mixed findings of studies that explored the influence of coursework on graduates' instructional quality, those that examined graduates' self-perceived preparedness found much more positive results. Ronfeldt et al. [46] described that completing more coursework before student teaching was positively associated with graduates' feelings of preparedness. Still, this coursework was unrelated to their cooperating teachers' evaluations of candidates' preparedness and, notably, negatively related to graduates' first-year observation ratings.

Research finds both positive and negative contributions of student teaching [34,47]. In the US, student teaching frequently lasts one semester, 10–16 weeks. The practical experience student teachers gain during student teaching may serve to develop their pedagogical content knowledge further, particularly because of the amount of time they have complete responsibility for teaching a class and stretching their content knowledge, but the effectiveness of this experience is strongly influenced by the nature of the school placement and the quality of the mentoring that student teachers receive during this experience [34]. At the same time, there is evidence that the overall length of the student teaching experience does not significantly increase elementary teachers' mathematical knowledge for teaching [48]. Further, Ronfeldt and Reininger [49] found that the duration of student teaching is unrelated to various teacher outcomes, including self-efficacy, feelings of preparedness, and plans to remain in teaching.

2.2. Teaching Quality Understood as Instructional Rigor

The Common Core State Standards in Mathematics [7] and the Next-Generation Science Standards [8] share a goal of K-12 students developing proficiency in the discipline. K-12 students should develop a deep understanding of key ideas by engaging in disciplinary practices, and both posit that such proficiency is gained as students engage in disciplinary practices. However, problem-solving and making sense of puzzling phenomena require a degree of rigor in student thought, which is not widely the norm in K-12 settings in the US. If teachers are to engage their students in rigorous instruction aligned with the reforms [11], shifts in the way science and math are currently taught will be required [12], and these shifts will require teachers well equipped for such work. Much of the literature discussed in the previous section used some measure of students' success on state-administered standardized tests as a proxy for instructional quality. However, as Ronfeldt [34] described, such measures may not capture the depth of learning that these reforms describe. For this reason, he calls for more studies linking preparation features to teaching quality measures beyond students' standardized test scores or self-perceptions of preparedness. Several studies in math and science education approach teaching quality by describing the rigor of a teacher's instruction [50–52] and the conceptualization employed in this work.

Researchers have often relied on classroom observations to explore the quality of teachers' instruction. This technique can be very insightful [53] but can also be particularly fraught with difficulties, as researchers need to be physically present in classrooms or record those classes [54]. Other researchers have advocated for closely examining tasks students are assigned to complete in science and math classrooms [55–58]. **Curricular tasks** are classroom-based activities that intellectually engage students with science content and/or practices [56,59]. Tasks not only shape the substance of what students learn but also how students think about, make sense of, and engage with the subject matter [58,60]. Thus, tasks are an essential tool for understanding the sorts of opportunities afforded to students, and different tasks require different levels and kinds of student thinking [56–59]. Tasks can be analyzed using the Task Analysis Guide in Mathematics [61] and its counterpart, the Task Analysis Guide in Science (TAGS) [59], as a lens to document the various opportunities for demand on students' thinking; in other words, to measure the cognitive demand observed in design. However, not all tasks provide similar opportunities for students to engage with the discipline, and different tasks have different degrees of demand on student thinking [54,60,62].

The NGSS and CCSSI outline expectations for students' learning, and research has begun to focus attention on the characteristics of the specific curricular tasks that engage students in such learning [59,63,64]. Instruction and classroom activities that engage students in math and science disciplinary practices place a high cognitive demand on students' thinking, positioning them to make sense of puzzling science phenomena and math problems as they develop a deeper understanding of math and science content and facility with math and science [59,65–67]. Considering the vision of instruction portrayed in the CCSSI and NHSS, the quality of instruction employed in the research presented here is defined by **instructional rigor**; that is, the degree to which students engage in high levels of thinking and reasoning as they work on cognitively demanding science or math tasks that they are assigned.

However, designing rigorous tasks is not enough for quality instruction. Research has shown that teachers often lower the rigor of such tasks during implementation [63,68]. In response, the field has begun examining whether students are engaging in rigorous math and science instruction; that is, whether students are engaging in curricular tasks in ways that place and maintain demand on their thinking through examining the implementation of the task [54].

Tasks are now recognized to play an essential role in the **instructional core**—a comprehensive view of science instruction that requires attention to the task, teachers' practice, and students' engagement. In [69], through close examination of the same lesson employed across three different classrooms, we showed how a teacher's instructional practices interacted with students' intellectual engagement with a cognitively demanding task to support the rigor of instruction. Research that informs our understanding of the role of tasks in instructional core in both math and science includes that of Lee and colleagues [70], who examined the work of 54 preservice elementary teachers in identifying the cognitive demand of tasks, increasing the demand of low-demand tasks, and anticipating student responses to these modified tasks. They found that preservice teachers were more successful in identifying low-demand tasks than those of higher demand, and participants were not always successful in increasing the demand of tasks. However, they could increase the rigor appropriately if taught to recognize the essential components of higher-level tasks. They sometimes anticipated adverse student reactions to higher-demand tasks. A contrasting piece of research that informed our work was that of Tekkumru-Kisa and colleagues [71], who examined the intellectual demand of 225 tasks identified as rigorous by a statewide sample of 125 in-service teachers. When examined by researchers, it was found that these tasks were instead classified as requiring low levels of intellectual work by students.

Further, teachers' decision-making around selecting tasks revealed limited attention to the intellectual engagement required by the task. Taken together, these earlier efforts with both in-service and preservice teachers suggest that both science and math teachers are more familiar with low-rigor tasks in science and math, although the findings of both Lee et al. [70] and Tekkumru-Kisa et al. [71] suggest that teachers do sometimes focus on the intellectual demand of tasks. Additionally, Lee and colleagues [70] advised that preservice teachers can be supported to attend to the demand of tasks used in instruction. How novice math and science teachers orient themselves to and employ intellectually demanding tasks remains an open question.

2.3. Teacher Persistence

It has long been recognized that teachers may become more effective with experience [72,73]. Consequently, if we are to achieve the vision portrayed in the CCSI and NGSS, it will be essential to retain successful teachers engaging in rigorous instruction. The first five years of practice are crucial because this is when novice teachers are honing their skills [74,75]. Many supports, such as teacher preparation and induction programs, work to bolster teachers as they develop their practice and gain experience in the initial stages of their careers [76].

However, despite this support, Ingersoll and colleagues [77] found that forty percent of novice teachers leave the profession within this period. This rate of attrition is substantiated by others [22] and is disproportionately high in the disciplines of math and science [78], among teachers of color, and in high-needs settings [38,79]. Such a constant churn of teachers limits students' learning opportunities, particularly for minoritized learners [38,80]. This phenomenon has been exacerbated by recent upheavals due to the global pandemic, issues of school safety [81,82], and persistent inequities in education [83–85].

Much of the current work on teacher retention uses an organizational perspective to describe the economic impacts of teacher attrition using the terms stayers, leavers, and movers [86,87]:

(1) **Stayers:** describes current K-12 classroom teachers at a specific school.
(2) **Leavers:** describes K-12 teachers who have left the classroom.
(3) **Movers:** describes current K-12 teachers who move schools.

This focus on teacher retention has given rise to terms such as "teacher turnover", which describes the losses that movers and leavers have on the field of education [86,88]. However, the explicit focus on the educational "cost" of attrition negatively views teachers who are movers, regardless of their motivation to move. It also limits the field's ability to understand if leavers are remaining in education, making it difficult to understand what may be affecting teachers' persistence in the classroom and education overall.

Considering these realities, we drew from the work of Larkin and colleagues [89] to frame our work regarding teacher persistence. This reframing is an essential shift because teachers' trajectories are typically characterized by teacher mobility and attrition (e.g., [90,91]). Our view of persistence moves beyond the impact on the employer to one that focuses on the effects of a teacher's persistence in the broader education profession [92]. This focus is vital because teachers who move from the classroom into other positions in education (such as administrator, STEM coordinator, or curriculum designer) provide contributions crucial to our school systems and structures [93]. Thus, utilizing persistence allows us to move away from focusing on teacher attrition and, instead, understand what supports teachers to stay. This reframing is essential to understanding what can be done within teacher education programs to better support their graduates to remain in the field, and it is vitally important for supporting educational equity.

3. Research Questions

This research aimed to explore the characteristics of teacher preparation programs that support their graduates in engaging their students in high-quality instruction and identify the program characteristics that support teachers in persisting in the field. These findings can inform the design of math and science teacher preparation programs that will prepare and support these teachers in ways necessary to support all students' math and science learning and for these teachers to persist in the field. The specific research questions guiding this study are:

What is the relationship between teacher preparation program design features,

(1) *and the quality of novice teachers' math and science instruction (as measured by the rigor of their planning and enactment)?*
(2) *and their graduates' persistence in teaching?*

4. Methods

This research is structured as a quantitative, non-experimental study exploring the relationship between (1) math and science teachers' preparation program features and the instructional quality of their novice graduates' instruction (conceptualized here as instructional rigor), and (2) math and science teachers' preparation program features and their graduates' long-term persistence in the profession. In this effort, we examined five math and science teacher preparation programs from across the United States and collected data regarding teacher persistence and instructional rigor from 46 of their graduates. Data about the structure of each of these programs were collected. In addition, data were collected from graduates of these programs during two stages of their teaching careers—six sets of instructional tasks (including teacher notes, task statements, student work, and teachers' evaluation of that work) to provide insights into their instructional rigor employed during their early years in teaching, and self-reported data about their graduates' employment three to eight years into their teaching. A timeline of the data collected for each research question can be found in Figure 1.

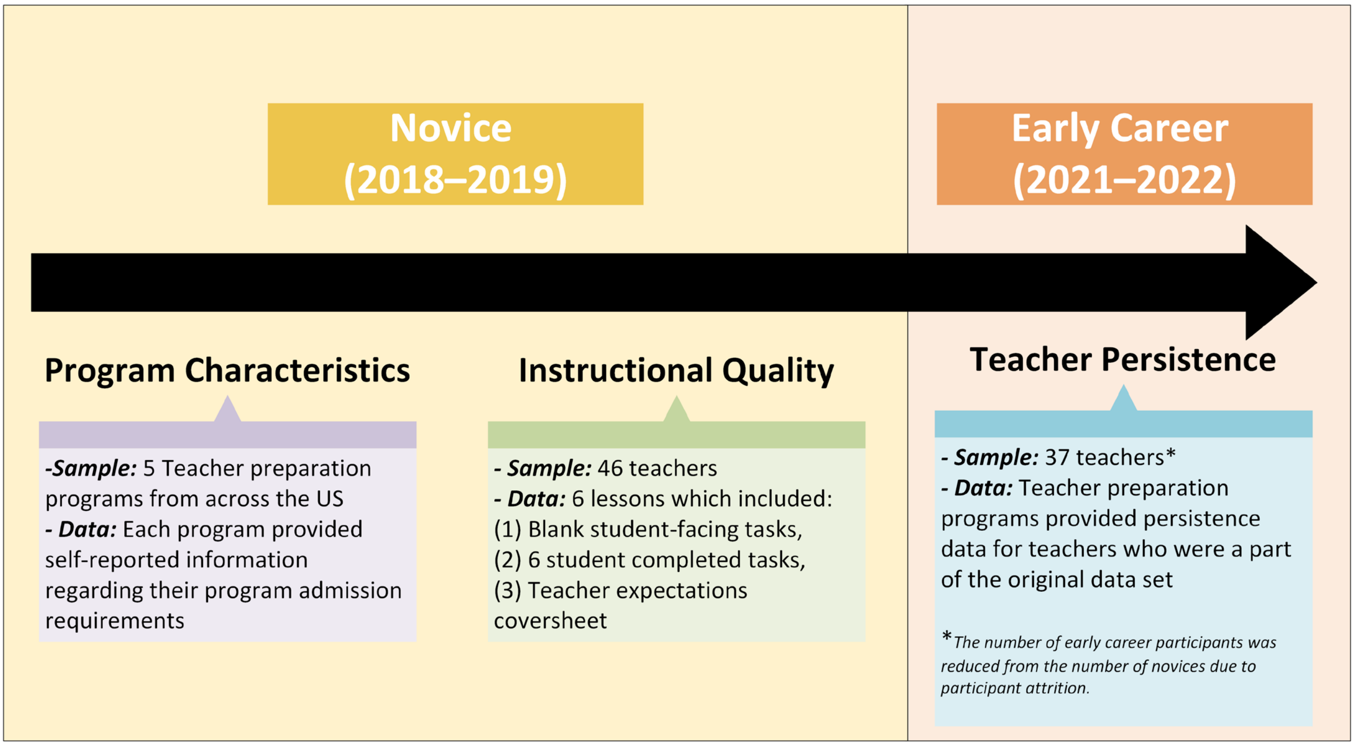

Figure 1. Program and teacher data collected across time.

4.1. Programs and Participants

Our data were drawn from Investigating Relationships between STEM Teacher Preparation, Instructional Quality, and Teacher Persistence (NSF) Grant, which included data from five math and/or science teacher preparation programs from different states across the United States [94]. Each teacher preparation program housed the Robert Noyce Teacher Scholarship Program, and background information on the program and participants can be found in Table 1. The Noyce Program is an ongoing, federally funded National Science Foundation project designed to foster the preparation and support of secondary math and science teachers across the US to teach in high-needs settings after graduation.

Table 1. Background information about teacher preparation programs and study participants.

Program Location			**Teachers**		
	Reform-Based Math Standards (i.e., Common Core)	**Reform-Based Science Standards (i.e., NGSS)**	**Math**	**Science**	**Total**
East Coast	Adapted	No	5	0	5
Intermountain West	Yes	Adapted	7	10	17
Northeast	Yes	Adapted	3	3	6
Southeast	Adapted	Adapted	1	2	3
West Coast	Yes	Yes	4	11	15 *
			Total: 46		

* Persistence data were available for 6 of the 15 teachers. Note: background information for the programs and teachers included in the study is presented.

Each math or science teacher who was a part of our study had completed their Noyce service requirement (two years teaching in a high-needs setting after graduation) and continued to teach beyond that period (between three and eight years). We collected data from each participant at two different time points: (1) novice years (zero to two years of teaching) and (2) early career (three to seven years of teaching).

During the 2018–2019 school year, each teacher preparation program contacted graduates in their novice years and asked them to participate in the study. A total of 46 teachers from 5 different programs participated in the study. After our initial participant data collection, focusing on their instructional quality, we followed up with these same participants four years later to collect data regarding their employment. These 5 programs produced more teachers, but only 46 agreed to participate in the study and/or completed data submission. Considering the complexity of the first set of data collection during their early years of teaching (submitting and commenting on six sets of classroom tasks with associated student work), our sample was skewed toward teachers who could select, organize, and submit this work. Thus, while we sought to include a wide range of math and science teachers, considering the voluntary and complex nature of the initial data collection, our sample included only novice teachers willing and capable of taking on this additional responsibility.

4.2. Data Sources and Analysis

4.2.1. Teacher Preparation Program Structure

To understand the role that teacher preparation programs have in instructional rigor and teachers' persistence, we identified the design characteristics of the five teacher preparation programs. The data used to understand the design characteristics were provided via an open-ended survey completed by the university's Noyce Program director. It included information about the overall university, its teacher education program, and its individual Noyce Program. We used literature in the field to identify what aspects of the program structure might influence teacher quality and persistence. In response, we focused on degree level (i.e., bachelor or post-bachelors), field hours, and coursework requirements, focusing on pedagogy-specific courses versus discipline-specific courses and the program foci [29,30,34,44] (see Appendix A for questions used in this survey).

4.2.2. Instructional Quality

Instructional Quality Data

To understand how participating teachers in their first one to three years of teaching conceptualized and implemented quality instruction using the lens of instructional rigor, in this work, we employed the IQA Mathematics Assignment Rubric, part of the IQA toolkit, which is a validated measure to assess instruction in math and reading/language arts through lesson observations and collections of student work [65]. For science tasks, we employed the Instructional Quality Assessment Science Assignment Rubrics (IQA-SAR), which include both observation and assignment rubrics built using Tekkumru-Kisa and colleagues' [68] Task Analysis Guide for Science (TAGS) framework. For this study, we utilized the assignment rubrics using student-facing work to provide a lens through which to understand what level and type of thinking students engaged in during the task. We utilized the assignment rubrics for the larger project because video data collected from classrooms can often be even more challenging to collect. After all, our programs (and their graduates) were spread across the country (with each school district having different requirements for collecting video data). At the same time, classroom work products are often much more widely available. It is important to note that the IQA and IQA-SAR rubrics are widely used, validated measures of instructional rigor in math and science, with an overall exact scale-point agreement between raters at 82% for math [65] and 76.9% for science [95]. The use of these rubrics to analyze instructional tasks, student work, and teacher evaluation of that work has been found to offer very similar results as those produced in classroom observations [65].

Both the IQA and IQA-SAR allow for the identification of the instructional quality of a task, as designed and implemented, to be assessed across different aspects of a task:

(1) The potential of rigor as designed.
(2) The rigor with which the task was implemented (as the task was assigned and assessed).
(3) The rigor is found in the kind and level of thinking teachers expect of their students.

By comparing the rigor scores across these three rubrics, researchers can gain insights into the teachers' levels of rigor and whether they could maintain the rigor of the task as designed during the implementation of the task. Each graduate in our dataset was asked to submit six tasks they had used in their classrooms that their students found challenging.

As seen in Table 2, each task included:

1. cover sheet that described the teacher's goals and expectations for students learning from the task,
2. The assigned task (or the student-facing work),
3. Six examples of student work on the task, along with the teacher's assessment of the quality of their work (including two from students who excelled at the task, two who adequately completed the task, and two who underperformed on the task) and teacher evaluations and comments on the student work.

These six tasks were collected at three-time points across the school year (two collected early in the school year, two mid-year, and two at the end).

Table 2. Description of IQA Mathematics and IQA-SAR Science Rubrics and data examined.

Rubric	Description	Data
R-1: Potential of the Task	Explores the potential of a task for engaging students in different kinds and levels of thinking in science or math, allowing for differentiation between tasks regarding the disciplinary activities that students engage in.	Instructional task assigned to students (i.e., worksheets and problems in texts).
R-2: Implementation of the Task	Explores the level and kind of thinking a majority of the students engaged in as they completed the task, highlighting instructional factors that maintain or reduce students' thinking throughout implementation.	Artifacts of students' work on the assigned task (differentiated by the teacher—2 low, 2 medium, and 2 high).
R-3: Teacher's Expectations	Explores the degree of rigorous thinking that science and math teachers expect from students throughout the lesson and in assignments.	Coversheet, which included teachers' directions to students and grading expectations, including any rubric they shared with the students.

Note: Qualitative descriptions of the IQA Mathematics and IQA-SAR Science Rubrics and the data analyzed are presented. Adapted from [59,65].

Instructional Quality Data Analysis

The task sets were analyzed using the Instructional Quality Assessment (IQA) Assignment Rubric for Mathematics, the design and validity of which were described by Boston, and the IQA Science Assignment Rubrics for Rigor (IQA-SAR), the design and validity of which were described by Tekkumru-Kisa et al. [54]. IQA assignment rubrics in math and science allow for the analysis of tasks designed by the teachers, the expectations the teachers provided for their students, and the associated student work to assess the quality of instruction, focusing on the rigor in students' opportunities for thinking and sensemaking. As shown in Table 2, we used three rubrics from both the IQA and IQA-SAR:

- **Rubric 1, or Potential of the Task**: describes the potential demand of the task on students' thinking (R-1 Potential).
- **Rubric 2, or Implementation of the Task**: describes the level and kind of thinking required from students as they complete the assigned task (R-2 Implementation).

- **Rubric 3, or Teacher's Expectations**: describes the degree of rigorous thinking teachers expect from students as they complete the task (R-3 Expectations).

Together, these rubrics structure a description of the quality of instruction with respect to rigor in students' opportunities to learn math or science around the task (for the entire set of IQA Mathematics Rubrics, contact Melissa Boston at bostonm@duq.edu, and for the IQA-SAR Science Rubrics, contact Miray Tekkumru Kisa at mtkisa@rand.org).

Using these rubrics, the data were categorized into one of four categories of academic rigor, ranging from:

- **Absent** (no math or science activity required)—students are not engaged in math or science activity.
- **Low**—students use memorized procedures, formulas, or definitions, or where students are engaged in a preset procedure to arrive at an answer.
- **Moderate**—students are asked to engage in complex thinking but are not expected to describe it.
- **High**—students engage in exploring and understanding the nature of math or science.

For a description of the specific descriptor of the level of rigor used in these rubrics, see Table 3 (Rubric 1), Table 4 (Rubric 2), and Table 5 (Rubric 6). See Appendix B for an example of a math task set and its associated category of rigor codes and Appendix C for an example of a science task set and its associated category of rigor.

Table 3. Descriptors of categories of rigor for the potential for academic rigor in the task (R1).

Rubric and Data	Category of Rigor	Math Descriptor	Science Descriptor
R1 Academic Rigor in Task Potential (Task)	High	**Problem Solving:** The task has the potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships (that is, using complex, non-algorithmic thinking), and students must provide their reasoning. The task suggests no approach or pathway; thus, it requires wrestling with ambiguity for its resolution.	**Figuring Out:** The task has the potential to engage students in sensible versions of the actual intellectual work of science—requiring students to develop explanations through the use of three dimensions of science (disciplinary core ideas, crosscutting concepts, and scientific practices). The task requires wrestling with ambiguity to create this explanation.
	Moderate	**Apply Procedures:** The task has the potential to engage students in complex thinking (such as finding relations, analyzing information, and generalizing to a broader idea), but the task does not ask for students' reasoning. The emphasis is on following a prescribed procedure for sensemaking but without an explanation of reasoning.	**Learning About:** The task has the potential to engage students in complex thinking and high-level cognitive processes (such as finding relations, analyzing information, and generalizing to a broader idea), but science content or scientific practices are forefronted. The emphasis is on learning about science content or practices.
	Low	**Rote/Procedural:** The potential of the task is limited: either to engaging students in using a specified procedure, or its use is evident or engages students in reproducing memorized information (facts, rules, formulas, or definitions.)	**Rote/Procedural:** The potential of the task is limited: either to engaging students in using a specified procedure, or its use is evident or engages students in reproducing memorized information (facts, rules, formulas, or definitions).
	Absent	**Absent:** The task requires no mathematical activity.	**Absent:** The task requires no scientific activity.

Note: potential for academic rigor in the task (R1) descriptors for math and science, adapted from [59,65].

To characterize the instructional quality undertaken by these math and science teachers during their novice years, using these rubrics, coders analyzed 209 tasks, including 81 math tasks and 128 science tasks, from teachers from all 5 institutions. For both the math and science tasks, six coders were trained in the use of the rubrics to score math and science tasks. Coders worked together to score example tasks with each of the three rubrics for math, coming together to discuss their ratings and rationales until they had achieved an inter-rater reliability of ~90% in their scores. After achieving an inter-rater reliability of ~90% in their scores for each of the three rubrics on these example tasks, they individually scored task sets drawn from this research. They completed the same process for the science rubrics and tasks.

Table 4. Descriptors of categories of rigor for implementation of the task (R2).

Rubric and Data	Category of Rigor	Math Descriptor	Science Descriptor
R2 Academic Rigor of Task Implementation (Student Work)	High	**Problem Solving:** Students' work indicated that students were engaged in problem solving, as students were engaged in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships (that is, using complex, non-algorithmic thinking), and explained their reasoning. Variation in students' problem solving suggested that the procedures were not predetermined.	**Figuring Out:** Students' work indicated that students were engaged in sensible versions of the actual intellectual work of science, and students drew on explanations using three dimensions of science (disciplinary core ideas, crosscutting concepts, and scientific practices) to develop explanations. Variations in students' work indicated that students wrestled with ambiguity when creating this explanation.
	Moderate	**Apply Procedures:** Students' work indicated that students were engaged in problem solving, but students did not explain their reasoning. Uniformity of students' work suggests that the procedures were prescribed.	**Learning About:** Students' work indicated engagement in complex thinking, primarily focused on either science content or practices and an emphasis on knowing and understanding content or practices.
	Low	**Rote/Procedural:** Students' work indicated that they engaged with the task at a procedural level, applying prescribed procedures to provide the correct answer, showing the steps, or students reproduced memorized information (facts, rules, formulas, or definitions).	**Rote/Procedural:** Students' work indicated engagement in procedures that led them to complete the task without knowing (or needing to know) why and how the script led to that answer and indicated students' use of skills/mechanics associated with the practices, or students reproduced memorized information (facts, rules, formulas, or definitions).
	Absent	**Absent:** Students' work provided no evidence of mathematical activity.	**Absent:** Students' work provided no evidence of scientific activity

Note: academic rigor of task implementation (R2) descriptors for math and science, adapted from [65,68].

Table 5. Descriptors of categories of rigor of teacher expectations (R3).

Rubric and Data	Category of Rigor	Math Descriptor	Science Descriptor
R3 Academic Rigor of Teacher Expectations (Cover Sheet)	High	**Problem Solving:** The majority of the teachers' expectations were for students to engage with the high-level demands of the task, such as using complex thinking and/or exploring, and understanding mathematical concepts, procedures, and/or relationships.	**Figuring Out:** The majority of the teachers' expectations were for students to engage in sensemaking, using the SEP, DCI, and CCC together in the service of explaining (i.e., figuring out) a phenomenon (i.e., productive engagement in practices was indicated in the teachers' expectations).
	Moderate	**Apply Procedures:** The teacher expected students to engage in complex mathematical thinking, but scaffolds were provided that lessened the demand on student thinking.	**Learning About:** The majority of teacher expectations were for students to engage in complex thinking, but either science content or scientific practices were forefronted. The emphasis was on learning about content or practices.
	Low	**Rote/Procedural:** Teacher expectations focused on student learning but not on complex thinking (e.g., expecting the use of a specific problem-solving strategy, and expecting short answers based on memorized facts, rules, or formulas), or teacher expectations were not focused on math (i.e., following directions, neat work, and student effort).	**Rote/Procedural:** Teacher expectations focused on student learning, but not on complex thinking (e.g., correct use of the prescribed procedure), or teacher expectations were not focused on science activity (i.e., following directions, neat work, and student effort).
	Absent	**Absent:** No teacher expectations for student work were found.	**Absent:** No teacher expectations for student work were found.

Note: academic rigor of teacher expectations (R3) descriptors for math and science, adapted from [65,68].

4.2.3. Teacher Persistence

To understand teachers' longer-term persistence in their work, three to four years after their induction period, and thus five to seven years after graduation, participants provided self-reported data regarding each teacher's teaching/career choices to their program, who shared this information with us. Using an adapted version of Larkin and colleagues' [92] person-position framework, as seen in Table 6, which highlights distinctions between a person and their teaching position, programs contacted each graduate and queried them about their employment history (see Appendix D for the interview protocol employed to collect these data). By comparing their previous and future employment, the field can explore any changes or variations teachers may make in their employment out of and within the field of education. The person-position conceptual framework describes a teacher's employment status using three categories:

(1) **Active**: teachers in this category are current teachers in K-12 teaching positions.
(2) **Reserve**: Teachers are not currently teaching in K-12 settings and can be separated into those taking a break in service and those who have begun providing advancement to the education field. Teachers taking a break in service see the break as temporary and intend to return to the same or a similar position. Those teachers providing advancement to the education field have joined an education-related position for which teacher certification has value (e.g., administration, higher education, or informal education).
(3) **Attritted**: teachers leaving the profession with no intent to return.

Table 6. Teacher persistence descriptions.

Employment Status	Description
Active	Current K-12 teachers
Enhancers	Current K-12 teachers who have taken on additional responsibilities that enhance their school or district (e.g., PLC leader, curriculum designer, mentor teacher, etc.)
Advancers	Former K-12 teachers who are employed in education-related positions where teacher certification has value (e.g., administration, higher education, or informal education)
Attritted	Former K-12 teachers who are no longer employed in the educational profession

Note: description of employment status in relation to teacher persistence.

Using Ingersoll's previous economically focused framing [86], active teachers would have been categorized as stayers, leavers, or movers, depending on whether they remained in a district. However, using the term active allowed us to describe any current K-12 teacher, regardless of location. Moreover, all teachers who are a part of the reserve and attritted would have been described as leavers. However, teachers providing advancement to the field are a positive and necessary part of the field of education [96,97]. Thus, their inclusion allowed us to explore those teachers who followed a career trajectory that led them from active to advancement status.

We built on the action element of Larkin and colleagues' framework to better describe our active and reserve teachers. As seen in Table 6, teachers were separated into four categories of employment: (1) **Active**:, or current teachers in K-12 teaching positions; (2) **Enhancers**, or teachers who are current teachers in K-12 teaching positions but who take on additional tasks that enhance the profession (e.g., professional learning community leader, curriculum designer, mentor teacher, etc.); (3) **Advancers**, or reserve teachers who have joined an education-related position for which teacher certification has value and is advancing the instructional efforts for schools, districts, or the field of education; (4) **Attritted**, or teachers who are leaving the profession with no intent to return. To understand emerging trends, we separated our data by program and discipline (see Appendix D for the interview protocol employed to collect these data).

4.2.4. Statistical Analysis

As we approached the statistical analysis of the dataset, we needed to account for the unique nature of our data. The forty-six participants in our research were graduates from one of five programs, and the six data points used to generate descriptions of the graduates' instructional rigor were drawn from the same classroom and teacher. Specifically, we sought to compare the persistence and rigor of graduates of each program. To answer these questions, we accounted for the non-independence of teachers or programs in mixed-effect models fit in a Bayesian framework using the package *brms* [98] in statistical language R. This approach allowed us to model repeated data measured at the scale of programs or instructional rigor, taking into account nested observational structures.

There were very few observations in non-"active" teacher persistence categories, making it challenging to fit models predicting outcomes of each of the four teaching levels. To overcome this, "active" and "enhancer" categories were combined (into "more active"), and "advancer" and "leaver" were combined (into "less active"). We then fit mixed-effect models, asking (1) whether teachers were more or less active and (2) whether instructional rigor was based on hour breakdowns of each program (field hours, STEM hours, and education hours), discipline (science or math), and program degree level (undergraduate or graduate). We scaled hour breakdowns to have a mean of zero and a standard deviation of one for improved model fit. We assumed a binomial link function to fit logistic regression models to the activity level or cumulative probit link functions to rigor scores with Bayesian mixed-effect models in the package *brms* in R [98]. For the model predicting teacher activity, observations were recorded at the level of individual instructors, nested within programs, so we included program identity as a random effect. For models predicting instructional rigor, observations were recorded at the level of lessons nested within instructors, so we included teacher identity as a random effect. More complicated error structures in rigor models encountered model fit problems (e.g., teachers nested within programs as random effects), presumably due to low variation within programs and small sample sizes. We ran models with 4 chains, 2000–20,000 iterations per chain, and a burn-in of 1000–10,000 iterations, assuming default conservative priors in *brms*. We evaluated model fit using posterior predictive checks and verified no fit issues concerning divergent samples, chain convergence, or low effective sample sizes of posteriors. We reported, in general, the 95% and 80% credible intervals (CIs) of estimated relationships and encouraged weighing interpretations of evidence of relationships at both more conservative (95%) and less conservative (80%) scales, considering the limited available data.

Because of low variation in hour breakdowns between undergraduate and graduate programs, many degree-level differences were colinear with program field, discipline-specific, and education hours. This introduced potentially erroneous correlations that we address below. Still, limited sample sizes and variation hindered our ability to account for differences in degree level or hour breakdowns when interpreting the opposing predictor. To interrogate these patterns, we fit secondary models pitting degree level and one of each field, STEM, or education hours in three separate models.

5. Findings

5.1. Teacher Preparation Program Characteristics

In the US, teacher education programs are approved by each state, and to be eligible to host a Noyce Program, the institution must be a state-approved secondary math and/or science teacher preparation program whose graduates are eligible for teacher licensure or certification upon graduation. Although the programs in our sample were drawn from distinct parts of the US, there were some commonalities across their requirements because they each included a Noyce Program. Overall, the programs in our study had similarities and differences in structure and focus, as seen in Table 7. Regarding degree level, our data included two undergraduate and three postgraduate programs. Postgraduate programs included two post-baccalaureate programs and one Master of Arts in Teaching (MAT) program.

Table 7. Teacher preparation program structure and focus.

Program Location	Degree Level	Requirements	Focus
East Coast *	Undergraduate	- 3.0 GPA - Working on a Math degree	Rigorous instruction
Intermountain West	Undergraduate	- 3.0 GPA - Working on a Math or Science degree	Culturally relevant pedagogy/Culture of care
Northeast	Master of Arts in Teaching (MAT)	- 3.0 GPA - M/S degree (120 h)	Culturally relevant pedagogy
Southeast	Post-baccalaureate	- 2.75 GPA - Content courses suggested (23 h)	None listed
West Coast	Post-baccalaureate	- 2.67 GPA - M/S degree (120 h)	Culturally relevant pedagogy/Social justice

* Math program only. M/S = math or science. Note: information regarding each teacher preparation program's degree level, requirements, and focus is presented.

Most undergraduate teacher education programs in the United States require a minimum of a 2.5 GPA for admission into an undergraduate program, typically beginning in the junior year of one's course of studies. Teacher preparation programs at the graduate level often require a 2.5 GPA or greater on undergraduate coursework as a pre-requisite for admissions. These same expectations were reflected in the admissions requirements for the five programs in our study. Thus, each state's requirements were very close in nature, limiting the variability in terms of admissions expectations for the five programs in our sample. Regardless of the program structure, students were required to have between a 2.67 and a 3.0 grade point average to gain admission to the program. Both undergraduate programs (those in the Intermountain West and East Coasts) required students to obtain undergraduate degrees in their discipline while obtaining the credits necessary to obtain a teaching certificate. Postgraduate programs in our sample ranged from requiring a degree in math or science (programs in the West Coast and Northeast) to just requiring 23 h of content courses (the program in the Southeast).

Regarding the programs' focus, programs in the West Coast, Intermountain West, and Northeast emphasized the importance of their graduates implementing culturally relevant pedagogy. The program on the West Coast also emphasized that graduates should become activists within their community to support social justice. The program in the Intermountain West emphasized that graduates should create a classroom culture that is caring and safe for all students, regardless of their background. The program on the East Coast emphasized that their graduates should develop proficiency in higher-level math so that they can then provide rigorous math instruction for their students. In contrast, the program from the Southeast had no overarching focus.

Regarding each program's characteristics highlighted, as shown in Table 8, we found that regardless of the degree level or program's structure, first, teachers were required to have gained classroom experience, requiring a minimum of 200 or more hours in the field. However, there were large differences in the required field hours and how they were distributed. The programs in the Intermountain West and West Coast required a total of 800 h, including 200 h of early fieldwork and a semester-long full-time student teaching experience. The East Coast and Northeast programs required fewer total hours, between 540 and 614, including slightly fewer early field hours and 2.5 to 3 months of student teaching. The most distinct program in our sample in terms of required field hours, the Southeast program, required 200 h of early fieldwork (including 50 h working as a teaching assistant in a course offered in the university system), after which point the students were hired as a full-time teacher of record, meeting regularly with a mentor; thus, the student teaching experience was omitted. This structure, seen in more and more programs in the United States, is in response to the acute teacher shortage many states are experiencing.

Table 8. Teacher preparation program characteristics overview.

Program Location	Degree Level	Field Hours (Early Field and Student Teaching)		Education Course Hours	Math and Science Content Course Hours *
		EF	ST		
East Coast *	Undergraduate	150	465	42	48
Intermountain West	Undergraduate	160	640	33	90
Northeast	Postgraduate	90	450	36	Math and science degree required for admission
Southeast	Postgraduate	200	0	39	23 h of math and science recommended for admission
West Coast	Postgraduate	160	640	37	Math and science degree required for admission

* Math program only. Note: information regarding each teacher preparation program's degree level, requirements for field hours, education coursework, and math and/or science content courses is presented.

In addition, each program required its graduates to complete coursework focused on how to teach. There was more uniformity here, with the hours ranging between 30 and 44, something to be expected, as these programs were all state-approved. However, there were also stark differences between undergraduate and postgraduate programs regarding the number of hours spent obtaining content-specific knowledge (i.e., microbiology, geometry, etc.). Postgraduate programs (those in our samples from the West Coast, Southeast, and Northeast) did not require their students to take any content courses during their program of study. However, it is important to note that a major in the content was required as a part of their undergraduate education to obtain admission to each program. The only deviation from this norm was the "requirement" for the program in the Southeast, which had the most lenient content requirements of any of the programs in our sample (only "suggesting" applicants complete 23 content-focused hours for admission). The structure of the post-baccalaureate/MAT programs, however, allowed for students' sole focus to be on math or science teaching during their studies.

5.2. Teacher Preparation Program and Instructional Rigor

Regardless of the degree level or program features, in terms of rigor, the five programs in this study produced teachers capable of quality math and science instruction. This is shown in Table 9, in the range data for the programs, where we see that each program was seen to have graduates capable of designing (R1) and implementing instruction (R2), whose ranges included moderate to high rigor levels.

The math teacher education programs in our sample, except for the program located in the Southeast, were seen to support their teachers to select/design tasks and implement them across a range of rigor, from low levels, where students were engaged in rote/procedural learning focused on following step-by-step instructions to answer problems, to the highest level of rigor, where students engaged in using their current math knowledge and problem-solving skills to answer a problem with multiple solutions. In contrast, the program from the Southeast was seen to support their math teachers to select/design tasks and implement them at a low level (rote/procedural) to a moderate level of rigor, where students were tasked with applying procedures to solve a problem. Notably, the Southeast program was the one that required the lowest number of field hours and math and science courses in our sample, suggesting a possible relationship between these factors.

Table 9. Teacher preparation program graduates' categories of instructional rigor.

Preparation Program		**Instructional Quality (Rigor Categories)**					
State	**Degree Level**	**R1-Potential**		**R2-Implementation**		**R3-Expectations**	
		Mode	**Range**	**Mode**	**Range**	**Mode**	**Range**
Math							
East Coast *	Undergraduate	Low	Low–High	Low	Low–High	Low	Absent–High
Intermountain West	Undergraduate	Low	Low–High	Low	Low–High	Low	Absent–High
Northeast	Postgraduate	Low	Low–High	Low	Low–High	Low	Low–High
Southeast	Postgraduate	Low	Low–Moderate	Low	Low–Moderate	Low	Low–Moderate
West Coast	Postgraduate	High	Low–High	Low	Low–High	Low	Low–High
Science							
Intermountain West	Undergraduate	Moderate	Absent–High	Low	Absent–High	Moderate	Absent–High
Northeast	Postgraduate	Moderate	Low–High	Moderate	Low–Moderate	Moderate	Low–High
Southeast	Postgraduate	Moderate	Low–High	Low	Low–High	Moderate	Low–Moderate
West Coast	Postgraduate	Moderate	Low–High	Low	Low–High	Moderate	Low–High

* Math program only. Note: findings regarding instructional rigor for math and science teacher preparation program graduates are presented. The background color differentiates different rubrics.

When we looked at the rigor of the instructional tasks selected/designed and implemented most frequently by the novice teachers in our sample in terms of each program, we found that apart from the program in the West, all math programs in our sample produced novice teachers who typically designed and implemented tasks at a low level of rigor (0–2, rote/procedural). However, the data suggested that the program found in the West produced graduates who designed highly rigorous lessons (problem solving) but were seen to reduce the rigor to a low level (0–2, rote/procedural) during implementation. This program was at the high or moderately high end of the spectrum for field hours, education courses, and math and science content courses.

Looking at the science programs in our sample, we found that all the programs in our sample supported their graduates to design lessons with a range of rigor from low (0–2, rote/procedural) to high (4–5, doing science). However, only the graduates from programs in the Intermountain West and West Coast were seen to support this range of rigor throughout their expectations (R-3) and implementation (R-2) and, interestingly, these are the two programs that required the highest number of field hours of the five programs in our sample.

Interesting patterns occurred when looking at the rigor level seen most often at each institution. The program graduates from the Intermountain West, West Coast, and Southeast were seen to design at a moderate level, engaging students in learning about science concepts or practices (requiring a moderate degree of academic rigor). However, this rigor was typically reduced to a low level or rote/procedural during implementation. Graduates from the program in the Northeast (which required a moderate degree of field hours and education courses but a high number of math and/or science courses) were seen to maintain rigor throughout design and implementation at the moderate level of learning about, suggesting that this program was the most successful in supporting their graduates to design and implement moderately academically rigorous lessons. There were other programs requiring a similarly high number of math or science hours whose graduates did not achieve this degree of rigor. However, it is important to note that, unlike all the other science programs, the Northeast's program was not seen to produce graduates who could implement instruction at a high level of rigor (doing science).

5.3. Teacher Persistence

To answer our second research question, which focused on the teacher preparation program's role in supporting teachers' persistence, we separated the programs by discipline. However, comparing math and science did not reveal any notable differences; thus, we examined the data persistence for math and science graduates together. We also examined course structure in terms of degree level, pedagogy coursework, math or science coursework, and fieldwork. Examining the various structural features of teacher preparation programs, there was no strong influence of the examined features (number of hours of math or science coursework, pedagogy coursework, and field hours) in shaping their graduates' persistence. We only saw notable differences in their graduates' persistence when we compared programs by different degree levels (undergraduate or postgraduate). See Table 10 for data regarding teachers' persistence.

Table 10. Teacher preparation program's role in teachers' persistence.

Teacher Preparation Program	Teacher Persistence			
	Active	**Enhancers**	**Advancers**	**Attritted**
		Undergraduate		
Intermountain West $n = 17$	~71% (12)	~6% (1)	~12% (2)	~12% (2)
East Coast * $n = 5$	20% (1)	0% (0)	20% (1)	60% (3)
Undergraduate Total: n = 22	***59% (13)***	***5% (1)***	***14% (3)***	***23% (5)***
		Postgraduate		
West Coast ** $n = 6$	~25% (1) **	50% (4) **	25% (1) **	0% (0)
Northeast $n = 6$	100% (6)	0% (0)	0% (0)	0% (0)
Southeast $n = 3$	100% (3)	0% (0)	0% (0)	0% (0)
Graduate Total: n = 15	***~67% (10)***	***~27% (4)***	***~7% (1)***	***0% (0)***
Total: *n* = 37	**~62% (23)**	**~14% (5)**	**~11% (4)**	**~14% (5)**

* Math program only. ** Only provided data for 6 of the 15 teachers who participated. Note: findings regarding teachers' persistence for math and science teacher preparation program graduates are presented.

Looking across all programs regardless of the discipline focus, we found that 76% of the teachers in our study remained in the classroom (i.e., active and enhancers), with differences found across the structures of teacher education programs. Traditional undergraduate programs resulted in 64% persistence in the graduates, while postgraduate teacher education programs resulted in 93% persistence. However, that tells only part of the story of graduates' careers in education. When we compared teachers' persistence across the various teacher preparation programs, we found that all of the teachers (100%) who graduated from postgraduate programs remained in the field of education as active teachers, enhancers, or advancers. Traditional undergraduate programs were seen to support teachers in remaining in the field at a lower rate of 78% (i.e., active, enhancers, and advancers). However, when we compared degree levels, we found that undergraduate programs were seen to more often support teachers to become advancers (3, 17%) compared to graduate programs (1, 7%).

5.4. Modeling Influence of Program Structure on Teacher Persistence

Figure 2 summarizes the estimated relationships between teacher activity levels (more active vs. less active) and degree level, field hours, math and science course hours (STEM), education course hours, and discipline by presenting the median (black dot), 80% credible interval (thick line), and 95% credible interval (thin line) for science (green) and math teachers (blue), where upper facet labels denote the fixed-effect predictor structure of logistic regression models. Our findings show trained in undergraduate programs tended to have a lower predicted probability of being "more active" (falling into active or enhancer categories) than teachers trained in postgraduate programs (median predicted effect of undergraduate training: −3.6126479, 95% credible interval (CI): [−11.1386483, 0.9421589], and 80% CI: [−8.0150302, −0.9627383]), while the 95% credible interval (CI) overlap was zero for the relationship between undergraduate and postgraduate training, and the 80% CI fell beneath zero.

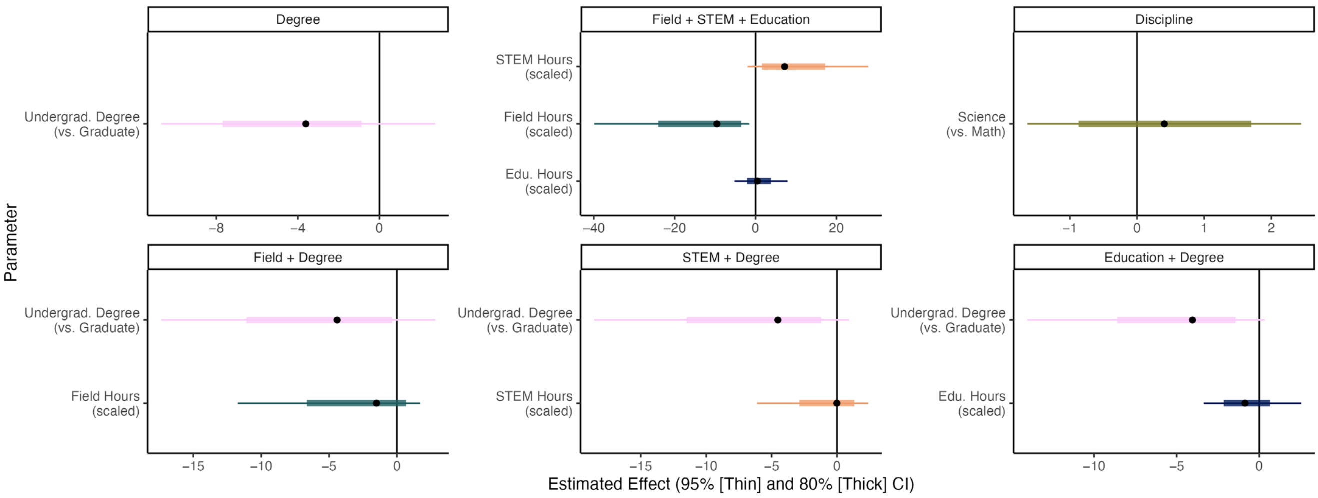

Figure 2. Comparison of the effect sizes (log-odds) of the relationships between varied program structures and the active/enhancer categories from logistic regression models.

Interestingly, as shown in Figure 2, we found that teachers trained in programs with more field hours tended to be less likely to be in the active/enhancer category (median: −9.45987877, 95% CI: [−38.62737084, −1.526906], and 80% CI: [−23.3895743, −3.509466]). In comparison, teachers from programs with more STEM hours tended to have slightly higher chances of being in the more active/enhancer category (median: 7.18482010, 95% CI: [−1.73223709, 27.217661], and 80% CI: [1.7086418, 16.933904]). However, this counterintuitive effect was likely due to collinearity between degree level and field hours (among other training components).

Not only did undergraduate programs tend to have fewer field hours, but in models directly comparing degree level and field hours, field hours lost evidence of relationships with the teacher activity level (median: −1.4892319, 95% CI: [−11.340355, 1.611557], and 80% CI: [−6.57391648, 0.6589338]), while degree level remained moderately negative (median: −4.3326749, 95% CI: [−17.793837, 2.665016], and 80% CI: [−11.00995236, −0.3648521]). See Figure 2 for this comparison.

The discipline of teachers (math or science) had little to no bearing on teacher persistence (median: 0.4458806, 95% CI: [−1.5868840, 2.4115472], and 80% CI: [−0.8893976, 1.71565655]).

5.5. Structure on Teachers' Instructional Rigor

We found very little evidence to suggest that the program features examined were strongly related to teachers' rigor scores (R1–R3), modeled using cumulative probit regression. We modeled rigor score independently (e.g., only modeling R1) by each discipline, math and science, as discipline scoring criteria were different and rigor types (R1–3) had different definitions and may not be assumed to be similar. In Figure 3, we summarized the estimated relationships between rigor scores and degree level as well as field hours, science and math hours (STEM), and education hours by presenting the median (black dot), 80% credible interval (thick line), and 95% credible interval (thin line) for science (green) and math teachers (blue), where upper facets denote the fixed-effect predictor structure of cumulative probit regression models. For math teachers, as shown in Figure 3, education hours were the only program predictor that had some (albeit very weak) evidence of a relationship with rigor (R1 and R2; R1—median: 3.959358×10^{-1}, 95% CI: [−0.25333960, 1.1711557], and 80% CI: [−0.01458919, 0.8478901]; R2—median: 3.309632×10^{-1}, 95% CI: [−0.40241940, 1.13316214], and 80% CI: [−0.1204412, 0.79867556]). For science teachers, as shown in Figure 3, education hours seemed to have a slightly reversed relationship, with again very weak evidence of a now negative relationship between education hours and R1 rigor scores (median: -2.466578×10^{-1}, 95% CI: [−0.6869442, 0.1605641], and 80% CI: [−0.51842149, 1.350817×10^{-2}]).

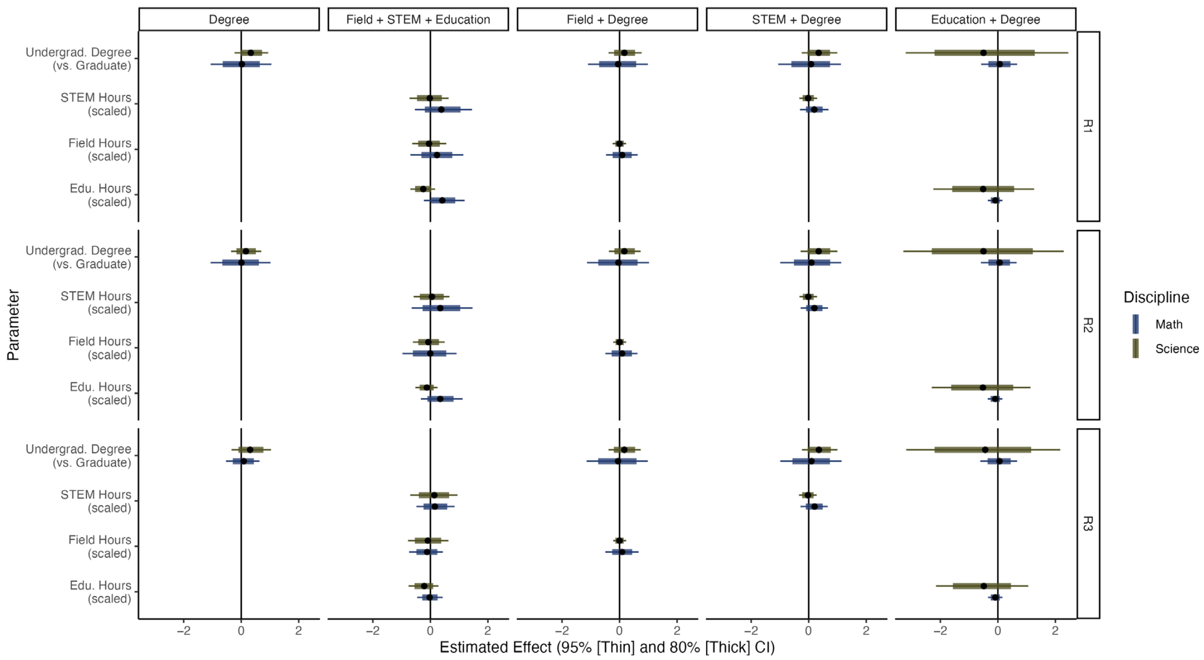

Figure 3. Comparison of the effect sizes of the relationships between varied program structures and math and science teachers' instructional rigor.

6. Discussion

It is widely recognized that if we are to produce students capable of using math and science to make sense of their worlds, as well as students who are interested in pursuing these disciplines as their careers, we will need to enhance the quality of the math and science learning opportunities that students are provided. Current reform efforts in math [7] and science [8] provide a vision for instruction that can challenge and engage students to think rigorously in math and science. Despite the proliferation of curricula informed by this vision, it is clear that teachers are a cornerstone of any effort to refine and hone what happens in classrooms [63,95]. Thus, our work identified the structures of teacher preparation programs associated with quality math and science teaching and teacher persistence in teaching.

To understand teaching quality in terms that are congruent with recent reform efforts, we broke away from the methods used in past work (which measured teaching quality through student success on standardized tests) to employ more direct observations of a teacher's instructional choices—using the lens of the academic rigor of a teacher's chosen tasks (described by R1), their evaluation of students' work on those tasks (described by R2), and their expectations of students' thinking on those tasks (described by R3).

Our findings suggested that teachers who graduated from each of the five teacher education programs in our sample were capable of designing and implementing rigorous math and science instruction. In this work, that capability was determined by the academic rigor of the task they selected or designed for their instruction, and our findings revealed that graduates from each of these programs were capable of selecting or constructing tasks of a high degree of rigor (see Table 9). This finding corroborates the common notion that how preservice teachers are prepared for the work of teaching influences their instructional approaches and their skills with those approaches [22–24].

We sought to address the ongoing ambiguity in the findings around the influence of program characteristics on teaching quality (understood here through the lens of the rigor of teachers' instruction). However, our findings, like those of other studies, revealed no strong link between program characteristics and the quality of instruction their graduates employed [30,34,48]. In our efforts to describe the differences between programs, we explored each program's home state's adoption of reform-based standards for math and science. We believed this might shed light on the differences in the level of rigor program graduates were seen designing and implementing. As seen in Table 1, although our five programs provided a range of adoption of reform-based standards for math and science, there was no clear link between standards' adaptation and/or adoption and the level and kind of rigor. Overall, our findings suggested that regardless of variability in the academics of the program (i.e., undergraduate vs. graduate), admission requirements, program focus, coursework, and field hours, the teacher preparation programs in our sample supported teachers to be able to design and implement rigorous lessons at a similar level of rigor.

Although the teachers in this study did show evidence of being able to design (R1), enact (R2), and have expectations of rigorous thinking (R3), it is important to note that such rigor was not the norm throughout the academic year for most of the teachers in the study. Indeed, when the teachers were asked to submit six tasks from throughout their school year that they believed were challenging for their students, the novice teachers in our sample, regardless of degree level or teaching context, most often provided lessons that were coded at the level of low to moderate rigor. Teachers' comfort with low-level tasks is similar to the findings of Tekkumru-Kisa [71] in their survey of the tasks understood to be rigorous by in-service science teachers, which in their analysis revealed that most of these tasks required low degrees of academic rigor on the part of students.

In our work, the mode of most math tasks designed and implemented by novice graduates of teacher preparation programs was coded at the level of low rigor for the potential of the task, implementation of the task, and teacher expectations, echoing the earlier work on math preservice teachers conducted by Lee and colleagues [70]. In contrast, novice teachers who graduated from science teacher preparation programs were most often seen to design and have expectations for their students at a moderate level of rigor. However, the data from students' work suggested that the teachers usually reduced the rigor during the task implementation. These trends align with the literature regarding the design and maintenance of rigorous instruction, showing that teachers often lower the rigor of tasks during implementation [63,68]. Concerning these findings, considering the need for all students to be engaged in rigorous math and science instruction, our data suggested that the programs in our study were successful in helping novice teachers design for rigor, but more attention is needed to support those same teachers in implementing those tasks in rigorous ways and holding high expectations for their students' intellectual efforts.

Regarding the relationship between different facets of the program structure and instructional quality (using the lens of intellectual rigor), we found no clear relationship between the numbers of pedagogical, discipline hours, field hours, and instructional rigor (see Figure 3). The findings regarding program characteristics and teachers' persistence were more promising. While our results revealed no influence of the examined features (number of hours of discipline-specific coursework, pedagogy coursework, and field hours) in shaping their graduates' persistence, only one program feature was influential. We found that only when we compared programs by different degree levels (undergraduate or postgraduate) did we see notable differences in their graduates' persistence in the classroom (see Figure 2). The result of our Multilevel Modeling suggested that the degree level of the program (i.e., undergraduate vs. postgraduate) may influence teachers' persistence in the classroom. Specifically, approximately 93% of teachers who graduated from a postgraduate program remained in the classroom (i.e., active or enhancers), in contrast to the 64% persistence rate for graduates of undergraduate programs. This finding contradicts the findings of other studies [22,29], which described that graduates of traditional teacher education programs (typically situated in bachelor programs) produce graduates with greater persistence. Our discrepant findings may be due to the narrow range of our post-bachelor programs, which, because of the requirement of state approval of programs, each of their requirements (field hours, pedagogy courses, and discipline-specific course requirements) were remarkably similar to the traditional programs. Thus, by selecting from state-approved programs, the range of our "alternative" sample may not have been discrepant enough to echo the findings of Marder [22] and others.

However, we believe, along with other scholars, that focusing our attention only on teachers currently serving in traditional roles in the classroom provides a limited view of persistence in education [92,96,97]. Considering this framing, it is essential to note that our data suggested that the teacher education programs in our sample create high-quality classroom teachers and the advancers necessary to support those teachers (as they accept positions as instructional coaches, curriculum designers, or STEM specialists).

Our findings revealed that graduate and undergraduate programs can support teachers in persisting and becoming advancers in their field. However, undergraduate teacher preparation programs in our sample were seen to graduate more advancers than graduate programs (14% vs. 7%). In related work [99], we found some evidence that teachers in our sample who teach highly rigorous instruction were more likely to move into advancer positions. This move could be because their instructional effectiveness was recognized by individuals in their school or larger district, or because they chose to go into higher education to address the deficiencies in their schools or broader educational system to better support other teachers to engage their students in high-quality instruction [99].

7. Limitations

It is important to highlight the influence of our sample on our findings. While we aimed to sample a wide range of teacher education programs in terms of their structure, because we sought to work with operating, state-approved teacher education programs, the sample we achieved only varied substantially regarding the number of field hours and degree level. The program sample may not have included enough variability to identify the influences of any one factor on teacher persistence or rigor.

Likewise, our methodological choices to measure instructional quality may have influenced our findings. We sought to use a more passive form of data collection to determine teacher quality—by collecting task sets, that is, materials teachers were already administering as part of their responsibilities. However, this requirement for participation in the study may have skewed our sample to those novice teachers capable of and willing to collect and organize this work for online submission. Thus, although we sought a more passive form of data collection, its organizational requirements served to select those capable of and willing to upload these task sets. Therefore, our high persistence numbers may have been due to the volunteer nature of our sample.

8. Conclusions and Future Research

This study compared graduates from five teacher education programs regarding the rigor of their instruction as it was designed and implemented during their novice teaching period and examined the persistence of these same novices years later. To understand rigor, we analyzed the tasks as designed, six samples of student work, and the expectations given to students, as described by the teacher, to determine the level of rigor observed in the design and implementation of the task. Our findings suggested, based on the similarities in terms of rigor supported in the design and implementation of tasks, that the five teacher education programs in our sample did not successfully support their graduates in designing and implementing highly rigorous tasks. However, all the programs produced graduates with more remarkable teacher persistence than that documented in other studies [17,20]. Based on these findings, we provide implications for research and practice in the following section.

First, though math and/or science teacher education programs may support novice teachers to be capable of designing and enacting rigorous tasks, our findings did not speak to any specific features of teacher education programs to be the factor that supports teachers to design and implement highly rigorous tasks consistently. The literature suggests that teachers have difficulty choosing rigorous tasks [54,62] and maintaining that demand throughout the implementation [63,99]. Considering that the findings presented here indicate that these novices could design for rigor, this rigor was inconsistent across their instruction. These findings suggest that the rigor seen in classroom task design and implementation is influenced by something other than the teacher preparation program. Our findings support those of Lee et al. [70], who described that the past experience of preservice teachers may well be oriented to lower-demand tasks, and they may not have experienced higher-level tasks. The finding that the science teachers had higher rigor in their tasks and expectations than that of their math colleagues calls for further research into the ways that organizational/district-level structures and norms (e.g., pacing guides, required curricula, required courses, etc.) and school-based norms (collaboration, perceived autonomy, etc.) shape teachers' ability to design and implement highly rigorous tasks.

Second, degree type was the factor that had the largest influence on teachers' persistence. However, our data did not allow us to explore this trend in greater detail. Thus, further research is needed to understand the differences between graduates from undergraduate and postgraduate teacher preparation programs to determine if the differences in teacher persistence are due to the nature of applicants for the different programs or the program's structure. These findings could then provide implications for what support might be necessary from undergraduate teacher preparation programs to ensure their graduates persist.

Although this work builds on the literature surrounding teacher attrition, we support a change in research to shift our focus from teacher attrition toward a focus on teacher persistence [89]. This change is crucial because instead of placing the focus on what is supporting teachers to leave, this framing allows researchers to examine the supports that enable teachers to stay in their field. We argue that understanding what supports math or science teachers to teach and ultimately persist in teaching [76,100], and to do so in ways that allow for rigorous instruction, is vital for educational equity.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci14050506/s1.

Author Contributions: Conceptualization, D.M.R. and S.A.S.; Methodology, D.M.R., W.R. and S.A.S.; Formal Analysis, D.M.R., W.R. and S.A.S.; Investigation, D.M.R. and S.A.S.; Resources, D.M.R. and S.A.S.; Data Curation, D.M.R. and S.A.S. Writing—Original Draft Preparation, D.M.R. and S.A.S.; Writing—Review & Editing, D.M.R., W.R. and S.A.S.; Visualization, D.M.R. and S.A.S.; Supervision, D.M.R. and S.A.S.; Project Administration, D.M.R. and S.A.S.; Funding Acquisition, S.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the National Science Foundation under #1661400. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Florida State University (IRB00000446, 03/06/2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and analyzed during the current study are not available because permission was not sought from participants to share the original data in this manner.

Acknowledgments: We want to thank the participants at five teacher preparation institutions and their novice teachers who were willing to invest the time and energy required by our data collection. We also want to thank Dr. Miray Tekkumru-Kisa for her insights in producing the instrumentation we employed to describe teachers' instructional rigor. Finally, the work would not have been possible without the efforts of our teacher coders, who were willing to come to understand and employ an instrument to examine rigor in teachers' instruction.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Characteristics of Preparation Program
Interview Protocol

a. What are the program admission requirements?
    i. GPA
    ii. GRE
    iii. ACT/SAT
b. What is the total number of hours required for graduation from the teacher preparation program?
c. What are the pathways in your teacher prep program? (MAT, undergraduate, career changer, etc.) What is the most common pathway in your program?
d. What are the general demographics of the student population in your program? (Age, ethnicity, sex, etc.)
e. What courses are requirements for subject matter knowledge?
f. What courses are requirements for pedagogical and other professional knowledge?
g. What is the main focus of each of these courses? If available, could we have a copy of a syllabus?
h. What courses are requirements for field and clinical experiences (i.e., field experiences, student teaching, mentoring, observations, PD, etc.)?
    i. What is the main focus of each of these courses? If available, could we have a copy of a syllabus?
    ii. Are any of these hours required to be in underserved settings? If so, how many?
i. Which of these courses touch on rigorous instruction (e.g., high-quality instruction/ambitious instruction)?
    i. Rigor is established by the opportunities for high-level thinking that is afforded to students situated in the task.
    ii. If they do, how many hours of the coursework focus on rigor/high-quality instruction/ambitious instruction? In what way do they have this focus?
j. Does your program offer any induction support?
    i. If so, please describe.
    ii. Were there any discussions regarding the suitability of schools as sites for first years of employment? Ways to negotiate parameters of employment (i.e., planning times/number of preps/club sponsors)?

Appendix B Example IQA- Categories for a Math Task Set

Rubric	Category of Academic Rigor	Score Descriptor	Coder's Comments
R1	Moderate	**Apply Procedures**: The task has the potential to engage students in complex thinking (such as finding relations, analyzing information, and generalizing to a broader idea), but the task does not ask for students' reasoning. The emphasis follows a prescribed procedure for sensemaking, but without explanation of reasoning.	The task asks students to create multiple representations, but not to explain the connection between them. In addition, students are asked to interpret which is the better deal, but are then led step-by-step on how to determine that.
R2	Low	**Rote/Procedural**: Students' work indicated that they engaged with the task at a procedural level, applying prescribed procedures to provide the correct answer, showing steps, or students reproduced memorized information (facts, rules, formulas, or definitions).	Most of the student work was uniform in nature and at the procedural level.
R3	Low	**Rote/Procedural**: Teachers' expectations focused on student learning, but not complex thinking (e.g., expecting use of a specific problem-solving strategy, or expecting short answers based on memorized facts, rules, or formulas), or teacher expectations were not focused on mathematics (i.e., following directions, neat work, and student effort).	Teachers' expectations focused on students' completion of work.

See Supplementary Materials.

Appendix C Example IQA-SAR Categories for a Science Task Set

Rubric	Category of Academic Rigor	Score Descriptor	Coder's Comments
R1	Moderate	Learning About: The task has the potential to engage students in complex thinking and high-level cognitive processes (such as finding relations, analyzing information, an dgeneralizing to a broader idea), but science content or scientific practices are forefronted. The emphasis is on learning about science content or practices.	Task asked students to use multiple representations to show their work and correct answers, but did not ask students to make connections between them.
R2	Low	Rote/Procedural: Students' work indicated engagement in procedures that led them to complete the task without knowing (or needing to know) why and how the script led to that answer, and indicated students' use of skills/mechanics associated with the practices or students reproduced memorized information (facts, rules, formulas, or definitions).	Majority of students used the same procedure following a template that the teacher said they discussed in class.
R3	Low	Rote/Procedural: Teachers' expectations focused on student learning, but not complex thinking (e.g., correct use of prescribed procedure), or teachers' expectations were not focused on science activity (i.e., following directions, neat work, and student effort).	Teacher expected students to use the specific procedure discussed in class to show their work and correct answers.

See Supplementary Materials.

Appendix D Interview Protocol for Teacher Persistence

- Personal
  - Why did you choose to become a teacher initially?
  - What are some of the things you enjoy or find satisfying about being a teacher?

    - What are some of the things about being a teacher that you don't enjoy or that you don't find satisfying?
    - Where did you complete your early years in teaching? Are you still there?
    - How long do you plan to continue to teach?
    - Have you considered leaving your school? The profession?
    - What would get you to stay? Or what prompted you to leave?
- Program
    - What role do you think your program played in the type of mathematics/science instruction you currently use with your students?
    - Also, what was your opinion of this image of ideal mathematics/science teaching that you observed?
    - What role do you think your program played in the fact that you have continued to remain a teacher/that you left teaching?
- Contextual factors
    - Why did you choose to teach in a high-needs school?
    - How long do you foresee teaching in a high-needs school?
    - What are your reasons for continuing to teach in/leaving a high-needs school?
    - What factors will influence how long you continue teaching in a high-needs school?
    - Was there one thing that has convinced you to stay teaching in a high-needs school?
    - Think about your current school environment. . .
        - Classroom autonomy is the freedom that teachers have in choosing textbooks, instructional techniques, classroom discipline, and grading policies. How would you describe the autonomy you were given in your classroom? (SLM)
        - How would you describe the administrative support that you were given at your school/schools? (SLM)
        - How would you describe the behavioral climate in your classroom? At your school? (SLM)
        - Describe what your relationships at school looked like with your students.
            - How well did you get to know your students?
                - If you knew them well, how did you build that knowledge/familiarity?

References

1. Aaronson, D.; Barrow, L.; Sander, W. Teachers and Student Achievement in the Chicago Public High Schools. *J. Labor Econ.* **2007**, *25*, 95–135. [CrossRef]
2. Rivkin, S.G.; Hanushek, E.A.; Kain, J.F. Teachers, Schools, and Academic Achievement. *Econometrica* **2005**, *73*, 417–458. [CrossRef]
3. Rockoff, J.E. The impact of individual teachers on student achievement: Evidence from panel data. *Am. Econ. Rev.* **2004**, *94*, 247–252. [CrossRef]
4. Desimone, L.M.; Smith, T.M.; Phillips, K.J. Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study of elementary students and teachers in Title I schools. *Teach. Coll. Rec.* **2013**, *115*, 1–46. [CrossRef]
5. Ekmekci, A.; Serrano, D.M. The Impact of Teacher Quality on Student Motivation, Achievement, and Persistence in Science and Mathematics. *Educ. Sci.* **2022**, *12*, 649. [CrossRef]
6. Hill, H.; Grossman, P. Learning from Teacher Observations: Challenges and Opportunities Posed by New Teacher Evaluation Systems. *Harv. Educ. Rev.* **2013**, *83*, 371–384. [CrossRef]
7. Council Of Chief State School Officers & National Governors' Association. Common Core State Standards Initiative. United States. 2011. Available online: https://www.loc.gov/item/lcwaN0010852/ (accessed on 3 January 2024).
8. National Research Council. *Next Generation Science Standards: For States, By States*; The National Academies Press: Washington, DC, USA, 2013. [CrossRef]
9. Ruiz-Primo, M.A.; Shavelson, R.J.; Hamilton, L.; Klein, S. On the evaluation of systemic science education reform: Searching for instructional sensitivity. *J. Res. Sci. Teach.* **2002**, *39*, 369–393. [CrossRef]
10. Cohen, D.K.; Ball, D.L. Policy and practice: An overview. *Educ. Eval. Policy Anal.* **1990**, *12*, 233–239. [CrossRef]
11. Banilower, E.R.; Smith, P.S.; Malzahn, K.A.; Plumley, C.L.; Gordon, E.M.; Hayes, M.L. *Report of the 2018 NSSME+*; Horizon Research, Inc.: Chapel Hill, NC, USA, 2018.
12. Reiser, B.J. What professional development strategies are needed for successful implementation of the Next Generation Science Standards. In Proceedings of the Invitational Research Symposium on Science Assessment; ETS Research Institute: Washington, DC, USA, 2013; pp. 1–22.

13. Marín Blanco, A.; Bostedt, G.; Michel-Schertges, D.; Wüllner, S. Studying teacher shortages: Theoretical perspectives and methodological approaches. *J. Pedagog. Res.* **2023**, *7*, 128–141. [CrossRef]
14. Oluk, S. Addressing the Teacher Exodus via Mobile Pedagogies: Strengthening the Professional Capacity of Second-Career Preservice Teachers through Online Communities of Practice. *Educ. Sci.* **2023**, *13*, 887. [CrossRef]
15. Nguyen, T.D.; Lam, C.B.; Bruno, P. *Is There a National Teacher Shortage? A Systematic Examination of Reports of Teacher Shortages in the United States*; Annenberg Institute for School Reform at Brown University: Providence, RI, USA, 2022.
16. Will, M. What will teacher shortages look like in 2024 and beyond: A research weighs in. *Education Week*, 21 December 2023.
17. Edwards, D.S.; Kraft, M.A.; Christian, A.; Candelaria, C.A. *Teacher Shortages: A Unifying Framework for Understanding and Predicting Vacancies*; Annenberg Institute for School Reform at Brown University: Providence, RI, USA, 2022; pp. 22–684.
18. Fuentes, S.Q.; Bloom, M.A. The Intricacies of the STEM Teacher Shortage. *Electron. J. Res. Sci. Math. Educ.* **2023**, *27*, i–vii.
19. Ingersoll, R.; May, H. *Recruitment, Retention and the Minority Teacher Shortage*; CPRE Research Report #RR-69; Consortium for Policy Research in Education: Philadelphia, PA, USA, 2011.
20. Ingersoll, R.; Perda, D. How high is teacher turnover and is it a problem. In *Proceedings of the Consortium for Policy Research in Education*; University of Pennsylvania: Philadelphia, PA, USA, 2010.
21. Grossman, P.; Loeb, S. *Taking Stock: An Examination of Alternative Certification*; Harvard Education Press: Cambridge, MA, USA, 2008.
22. Marder, M. Teacher shortages and the decline of secondary teacher production. In *Proceedings of the Annual Meeting of UTeach STEM Educators Association*; UTeach Institute: Austin, TX, USA, 2022.
23. Stroupe, D.; Gotwals, A.; Christensen, J.; Wray, K.A. Becoming Ambitious: How a Practice-based Methods Course and "Macroteaching" Shaped Beginning Teachers' Critical Pedagogical Discourses. *J. Sci. Teach. Educ.* **2022**, *6*, 683–702. [CrossRef]
24. Stroupe, D.; Hammerness, K.; McDonald, S. *Preparing Science Teachers through Practice-Based Teacher Education*; Harvard Education Press: Cambridge, MA, USA, 2020.
25. US Department of Education, Office of Postsecondary Education. *Meeting the Highly Qualified Teachers Challenge: The Secretary's Annual Report on Teacher Quality*; US Department of Education, Office of Postsecondary Education: Washington, DC, USA, 2005.
26. US Department of Education. *Higher Education Act Title II Reporting System*; Office of Postsecondary Education: 2019–2020.
27. US Department of Education. *National Teacher and Principal Survey (NTPS)*; Public School Teacher Data File; National Center for Education: 2020–2021.
28. Darling-Hammond, L. Keeping good teachers: Why it matters, what leaders can do. *Educ. Leadersh.* **2003**, *60*, 6–13.
29. Van Overschelde, J.P.; Wiggins, A.Y. Teacher Preparation Pathways: Differences in Program Selection and Teacher Retention. *Action Teach. Educ.* **2020**, *42*, 311–327. [CrossRef]
30. Stuart, J.S.; Tatto, M.T. Designs for initial teacher preparation programs: An international view. *Int. J. Educ. Res.* **2000**, *33*, 493–514. [CrossRef]
31. National Commission on Teaching & America's Future. *No Dream Denied: A Pledge to America's Children*; National Commission on Teaching and America's Future: Washington, DC, USA, 2003.
32. Kee, A.N. Feelings of Preparedness Among Alternatively Certified Teachers. *J. Teach. Educ.* **2012**, *63*, 23–38. [CrossRef]
33. O'Neill, S.; Stephenson, J. Does classroom management coursework influence pre-service teachers' perceived preparedness or confidence? *Teach. Teach. Educ.* **2012**, *28*, 1131–1143. [CrossRef]
34. Ronfeldt, M. *Links among Teacher Preparation, Retention, and Teaching Effectiveness. Evaluating and Improving Teacher Preparation Programs*; National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs: Washington, DC, USA, 2021. [CrossRef]
35. DeAngelis, K.J.; Wall, A.F.; Che, J. The Impact of Preservice Preparation and Early Career Support on Novice Teachers' Career Intentions and Decisions. *J. Teach. Educ.* **2013**, *64*, 338–355. [CrossRef]
36. Goldhaber, D.; Cowan, J. Excavating the Teacher Pipeline. *J. Teach. Educ.* **2014**, *65*, 449–462. [CrossRef]
37. Dove, M.K. Teacher attrition: A critical American and International education issue. *Delta Kappa Gamma Bull.* **2004**, *71*, 8–14.
38. Ingersoll, R.; Merrill, L.; May, H. *What are the Effects of Teacher Education and Preparation on Beginning Teacher Attrition?* Research Report [#RR-82]; Consortium for Policy Research in Education, University of Pennsylvania: Philadelphia, PA, USA, 2014.
39. Harris, D.N.; Sass, T.R. Teacher training, teacher quality and student achievement. *J. Public Econ.* **2011**, *95*, 798–812. [CrossRef]
40. Allen, M. *Eight Questions on Teacher Preparation: What Does the Research Say?* Education Commission of the States and the U.S. Department of Education's Office of Educational Research and Improvement: Denver, CO, USA, 2003.
41. Schmidt, W.H.; Cogan, L.; Houang, R. The role of opportunity to learn in teacher preparation: An international context. *J. Teach. Educ.* **2011**, *62*, 138–153. [CrossRef]
42. Boyd, D.J.; Grossman, P.L.; Lankford, H.; Loeb, S.; Wyckoff, J. Teacher preparation and student achievement. *Educ. Eval. Policy Anal.* **2009**, *31*, 416–440. [CrossRef]
43. Henry, G.T.; Campbell, S.L.; Thompson, C.L.; Patriarca, L.A.; Luterbach, K.J.; Lys, D.B.; Covington, V.M. The Predictive Validity of Measures of Teacher Candidate Programs and Performance. *J. Teach. Educ.* **2013**, *64*, 439–453. [CrossRef]
44. Preston, C. University-Based Teacher Preparation and Middle Grades Teacher Effectiveness. *J. Teach. Educ.* **2016**, *68*, 102–116. [CrossRef]
45. Kleickmann, T.; Richter, D.; Kunter, M.; Elsner, J.; Besser, M.; Krauss, S.; Baumert, J. Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *J. Teach. Educ.* **2013**, *64*, 90–106. [CrossRef]

46. Ronfeldt, M.; Matsko, K.K.; Nolan, H.G.; Reininger, M. Three Different Measures of Graduates' Instructional Readiness and the Features of Preservice Preparation That Predict Them. *J. Teach. Educ.* **2020**, *72*, 56–71. [CrossRef]
47. Nissim, Y.; Danial-Saad, A. The Resilient Teacher: Unveiling the Positive Impact of the Collaborative Practicum Model on Novice Teachers. *Educ. Sci.* **2023**, *13*, 1162. [CrossRef]
48. Youngs, P.; Qian, H. The influence of university courses and field experiences on Chinese elementary candidates' mathematical knowledge for teaching. *J. Teach. Educ.* **2013**, *64*, 244–261. [CrossRef]
49. Ronfeldt, M.; Reininger, M. More or better student teaching? *Teach. Teach. Educ.* **2012**, *28*, 1091–1106. [CrossRef]
50. Colley, C.; Windschitl, M. Rigor in elementary science students' discourse: The role of responsiveness and supportive conditions for talk. *Sci. Educ.* **2016**, *100*, 1009–1038. [CrossRef]
51. Thompson, J.; Hagenah, S.; Kang, H.; Stroupe, D.; Braaten, M.; Colley, C.; Windschitl, M. Rigor and responsiveness in classroom activity. *Teach. Coll. Rec.* **2016**, *118*, 1–58. [CrossRef]
52. Windschitl, M.; Thompson, J.; Braaten, M.; Stroupe, D. Proposing a core set of instructional practices and tools for teachers of science. *Sci. Educ.* **2012**, *96*, 878–903. [CrossRef]
53. Sofianidis, A.; Kallery, M. An Insight into Teachers' Classroom Practices: The Case of Secondary Education Science Teachers. *Educ. Sci.* **2021**, *11*, 583. [CrossRef]
54. Tekkumru-Kisa, M.; Kisa, Z.; Hiester, H. Intellectual work required of students in science classrooms: Students' opportunities to learn science. *Res. Sci. Educ.* **2021**, *51*, 1107–1121. [CrossRef]
55. Boston, M.D.; Smith, M.S. Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *J. Res. Math. Educ.* **2009**, *40*, 119–156.
56. Doyle, W. Academic work. *Rev. Educ. Res.* **1983**, *53*, 159–199. [CrossRef]
57. Hiebert, J.; Wearne, D. Instructional Tasks, Classroom Discourse, and Students' Learning in Second-Grade Arithmetic. *Am. Educ. Res. J.* **1993**, *30*, 393–425. [CrossRef]
58. Stein, M.K.; Grover, B.W.; Henningsen, M. Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *Am. Educ. Res. J.* **1996**, *33*, 455–488. [CrossRef]
59. Tekkumru-Kisa, M.; Stein, M.K.; Schunn, C. A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science. *J. Res. Sci. Teach.* **2015**, *52*, 659–685. [CrossRef]
60. Dempsey, M.; O'Shea, A. The role of task classification and design in curriculum making for preservice teachers of mathematics. *Curric. J.* **2020**, *31*, 436–453. [CrossRef]
61. Stein, M.K.; Smith, M.S.; Henningsen, M.A.; Silver, E.A. *Implementing Standards-Based Math Instruction: A Casebook for Professional Development*; Teachers College Press: New York, NY, USA, 2009.
62. Lee, J.-E.; Hwang, S.; Yeo, S. Preservice Teachers' Task Identification and Modification Related to Cognitive Demand. *Int. J. Sci. Math. Educ.* **2023**, *22*, 911–935. [CrossRef]
63. Kang, H.; Windschitl, M.; Stroupe, D.; Thompson, J. Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *J. Res. Sci. Teach.* **2016**, *53*, 1316–1340. [CrossRef]
64. Munter, C.; Correnti, R. Examining Relations between Mathematics Teachers' Instructional Vision and Knowledge and Change in Practice. *Am. J. Educ.* **2017**, *123*, 171–202. [CrossRef]
65. Boston, M. Assessing instructional quality in mathematics. *Elem. Sch. J.* **2012**, *113*, 76–104. [CrossRef]
66. Odden, T.O.B.; Russ, R.S. Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Sci. Educ.* **2019**, *103*, 187–205. [CrossRef]
67. Windschitl, M.; Calabrese Barton, A. (Eds.) *Rigor and Equity by Design: Locating a Set of Core Teaching Practices for the Science Education Community*; American Educational Research Association: Washington, DC, USA, 2016; pp. 1099–1158.
68. Tekkumru-Kisa, M.; Schunn, C.; Stein, M.K.; Reynolds, B. Change in thinking demands for students across the phases of a science task: An exploratory study. *Res. Sci. Educ.* **2019**, *49*, 859–883. [CrossRef]
69. Tekkumru-Kisa, M.; Ackil-Okan, O.; Kisa, Z.; Southerland, S. Exploring science teaching in interaction at the instructional core. *J. Res. Sci. Teach.* **2023**, *60*, 26–62. [CrossRef]
70. Lee, M.; Kim, J.W.; Mo, Y.; Walker, A.D. A review of professional learning community (PLC) instruments. *J. Educ. Adm.* **2022**, *60*, 262–287. [CrossRef]
71. Tekkumru-Kisa, M.; Preston, C.; Kisa, Z.; Oz, E.; Morgan, J. Assessing instructional quality in science in the era of ambitious reforms: A pilot study. *J. Res. Sci. Teach.* **2021**, *58*, 170–194. [CrossRef]
72. Darling-Hammond, L.; Wei, R.C.; Andree, A.; Richardson, N.; Orphanos, S. *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad*; Stanford University: Stanford, CA, USA, 2009.
73. Podolsky, A.; Kini, T.; Darling-Hammond, L. Does teaching experience increase teacher effectiveness? A review of US research. *J. Prof. Cap. Community* **2019**, *4*, 286–308. [CrossRef]
74. Darling-Hammond, L. President Obama and education: The possibility for dramatic improvements in teaching and learning. *Harv. Educ. Rev.* **2009**, *79*, 210–223. [CrossRef]
75. DeMonte, J. The Leaky Pipeline: Why Don't New Teachers Return? 2016. Available online: https://www.air.org/resource/blog-post/leaky-pipeline-why-dont-new-teachers-teach (accessed on 3 January 2024).
76. Wright, D.S.; Balgopal, M.M.; Sample McMeeking, L.B.; Weinberg, A.E. Developing resilient K-12 STEM teachers. *Adv. Dev. Hum. Resour.* **2019**, *21*, 16–34. [CrossRef]

77. Ingersoll, R.; May, H.; Collins, G.; Fletcher, T. *Trends in the Recruitment, Employment and Retention of Teachers from Under-Represented Racial-Ethnic Groups, 1987 to 2016*; Chapter in The AERA Handbook of Research on Teachers of Color; American Educational Research Association: Washington, DC, USA, 2021.
78. Walker, T. Teacher Shortage Is "Real and Growing, and Worse than We Thought. 2019. Available online: http://neatoday.org/2019/04/03/how-bad-is-the-teacher-shortage (accessed on 3 January 2024).
79. Boyd, D.; Grossman, P.; Lankford, H.; Loeb, S.; Wyckoff, J. *Who Leaves? Teacher Attrition and Student Achievement*; NBER Working Paper No. w14022; National Bureau of Economic Research: Cambridge, MA, USA, 2008.
80. Texas, E. *Texas Educator Preparation Pathways Study: Developing and Sustaining the Texas Educator Workforce*; University of Texas at Austin College of Education: Austin, TX, USA, 2022.
81. Pressley, T. Factors Contributing to Teacher Burnout During COVID-19. *Educ. Res.* **2021**, *50*, 325–327. [CrossRef]
82. Steiner, E.D.; Woo, A. *Job-Related Stress Threatens the Teacher Supply: Key Findings from the 2021 State of the U.S. Teacher Survey*; Rand Corperation: Santa Monica, CA, USA, 2021.
83. Bang, M.; Warren, B.; Rosebery, A.S.; Medin, D. Desettling expectations in science education. *Hum. Dev.* **2013**, *55*, 302–318. [CrossRef]
84. Madkins, T.C.; Morton, K. Disrupting Anti-Blackness with Young Learners in STEM: Strategies for Elementary Science and Mathematics Teacher Education. *Can. J. Sci. Math. Technol. Educ.* **2021**, *21*, 239–256. [CrossRef]
85. Morales-Doyle, D. Justice-centered science pedagogy: A catalyst for academic achievement and social transformation. *Sci. Educ.* **2017**, *101*, 1034–1060. [CrossRef]
86. Ingersoll, R. Is There Really a Teacher Shortage? 2003. Available online: https://www.gse.upenn.edu/pdf/rmi/Shortage-RMI-09-2003.pdf (accessed on 3 January 2024).
87. Kennedy, M.M. Attribution Error and the Quest for Teacher Quality. *Educ. Res.* **2010**, *39*, 591–598. [CrossRef]
88. Carver-Thomas, D.; Darling-Hammond, L. The trouble with teacher turnover: How teacher attrition affects students and schools. *Educ. Policy Anal. Arch.* **2019**, *27*, 1–32. [CrossRef]
89. Larkin, D.B.; Carletta, L.; Patzelt, S.P.; Ahmed, K. (Eds.) *Making Sense of Science Teacher Retention: Teacher Embeddedness and Its Implications for New Teacher Support*; American Association for the Advancement of Science: Washington, DC, USA, 2022; pp. 275–319.
90. Kukla-Acevedo, S. Leavers, movers, and stayers: The role of workplace conditions in teacher mobility decisions. *J. Educ. Res.* **2009**, *102*, 443–452. [CrossRef]
91. Palma-Vasquez, C.; Carrasco, D.; Tapia-Ladino, M. Teacher mobility: What is it, how is it measured and what factors determine it? A scoping review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2313. [CrossRef] [PubMed]
92. Larkin, D.B.; Patzelt, S.P.; Ahmed, K.M.; Carletta, L.; Gaynor, C.R. Portraying secondary science teacher retention with the person-position framework: An analysis of a state cohort of first-year science teachers. *J. Res. Sci. Teach.* **2022**, *59*, 1235–1273. [CrossRef]
93. Day, C.; Sammons, P.; Gorgen, K. *Successful School Leadership*; Education Development Trust: Reading, UK, 2020.
94. Travis, J.A.; Southerland, S.A.; Tekkumru-Kisa, M. Investigating Relationships between STEM Teacher Preparation, Instructional Quality, and Teacher Persistence. **August 2017–July 2023**, *1,399,709.00*.
95. Tekkumru-Kisa, M.; Stein, M.K.; Doyle, W. Theory and research on tasks revisited: Task as a context for students' thinking in the era of ambitious reforms in mathematics and science. *Educ. Res.* **2020**, *49*, 606–617. [CrossRef]
96. Cochran-Smith, M. The Problem of Teacher Education. *J. Teach. Educ.* **2004**, *55*, 295–299. [CrossRef]
97. Nieto, S. *What Keeps Teachers Going*; Teachers College Press: New York, NY, USA, 2003.
98. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **2017**, *80*, 1–28. [CrossRef]
99. Rhemer, D.; Schellinger, J.M.; Southerland, S. Examining the Interaction Between Preparation Programs, Instructional Rigor, and Math and Science Teachers' Persistence in Teaching. Presented at the 2023 AAAS Noyce Summit, Washington, DC, USA, 26–28 June 2023.
100. Sutcher, L.; Darling-Hammond, L.; Carver-Thomas, D. Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Educ. Policy Anal. Arch.* **2019**, *27*, 325–327. [CrossRef]