

Initial findings of the Physics Teacher Education Program Analysis rubric: What do thriving programs do?

Rachel E. Scherr¹ 

University of Washington, Bothell, Washington 98011, USA

Stephanie V. Chasteen² 

Chasteen Educational Consulting, UBC 390, Boulder, Colorado 80309, USA



(Received 24 August 2019; accepted 13 January 2020; published 3 April 2020)

Systematic research on physics teacher education (PTE) programs in the United States is rare, owing both to the great variety of practices and structures enacted by U.S. PTE programs and to the lack of measurement tools available to measure what successful PTE programs do. To help meet the need for a specific, objective, and reliable guide for research on PTE programs, the Physics Teacher Education Coalition developed the Physics Teacher Education Program Analysis (PTEPA) rubric, which characterizes the practices and structures observed at “thriving” PTE programs (defined as programs in the U.S. that frequently graduate five or more physics teachers per year). Initial research based on the PTEPA rubric suggests that thriving programs are strong in multiple areas (especially institutional commitment, leadership, and collaboration among partners in education and physics), and that several areas of strength align with those indicated by the existing PTE literature. However, thriving programs are not necessarily strong in all areas, instead reflecting local conditions at the institutional and state level. Such findings illustrate the opportunity for measurement and hypothesis testing about the most important features that PTE programs should have.

DOI: 10.1103/PhysRevPhysEducRes.16.010116

I. INTRODUCTION

Systematic research on physics teacher education (PTE) programs has the potential to guide program development. The Physics Teacher Education Coalition (PhysTEC) conducted a study of “thriving” PTE programs (defined as programs in the U.S. that frequently graduate five or more physics teachers per year) to identify common practices and structures of these programs [1]. By measuring what a variety of different U.S. programs do to support PTE, PhysTEC hopes to support others in creating or improving their own effective programs, thereby helping to address the severe shortage of high school physics teachers across the United States [2].

The thriving programs study led to the development of the Physics Teacher Education Program Analysis (PTEPA) rubric to characterize the practices and structures observed at thriving programs [3,4]. The opportunity that the PTEPA rubric provides for systematic research is significant. Existing research into U.S. PTE programs is valuable,

but mainly descriptive: it offers examples of what successful programs do, rather than empirical studies of these practices or their impacts on future teachers. The risk of learning from examples is that descriptive studies do not have the ability to differentiate between critical and noncritical features of a particular approach. A measurement tool such as the PTEPA rubric supports inquiry into which features of PTE programs are important for success, and in which contexts. The rubric is designed to characterize PTE programs and enable comparisons among programs, thus enabling measurement of program growth and allowing aggregation of data across different programs. Additionally, the rubric serves the goal of self-improvement, providing a specific, objective, and reliable guide for physics teacher educators seeking to improve their programs.

This paper describes what has been learned in initial application of the PTEPA rubric to eight thriving PTE programs. The primary research question is, “What do thriving PTE programs do?” The initial findings are as follows:

- Thriving programs are strong in multiple areas, and the pattern of strength varies across programs.
- The most consistent strengths among thriving programs are in “institutional commitment,” “leadership, and collaboration” among partners in education and physics.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](#). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

- C. Thriving programs are strong in some of the specific areas indicated by the existing PTE literature.
- D. Thriving programs are consistently strong in certain specific activities and structures.
- E. Thriving programs are not strong in all specific activities and structures.

Most of these findings had been suggested in earlier descriptive studies. The contribution of the present research is to offer a systematic approach to studying PTE, in which the practices and structures that comprise a PTE program are catalogued consistently along with defined levels of achievement for each element. Such systematization is crucial to an evidence-based argument: the PTEPA rubric's catalogue of standards, components, and items, each with a recommended level of accomplishment, presents the opportunity for measurement and hypothesis testing. With this instrument, researchers have the opportunity to discover correlations, notice patterns, and potentially challenge accepted wisdom.

II. LITERATURE REVIEW: WHAT SHOULD PHYSICS TEACHER EDUCATION PROGRAMS DO?

A world of literature addresses how best to prepare students to become effective teachers. The following brief review highlights the findings from this literature that are most relevant for physics-specific teacher education programs in the U.S., especially those features that concern disciplinary departments.

A. There is little systematic research on PTE programs

The National Research Council (NRC) report titled “Preparing Teachers: Building Evidence for Sound Policy” [5] indicates major gaps in our knowledge of what happens in U.S. teacher education programs, and what makes such programs effective. The charge for this effort was to answer the question, “What sorts of instruction and experiences do (U.S.) teacher candidates receive in preparation programs of various types?” The report states that “it was exceptionally difficult to assemble a clear picture of teacher preparation because there have been no systematic efforts to collect the necessary data; thus, we can provide only partial answers.” The report continues: “There has been an extraordinary amount of work, from a variety of fields, on questions about the factors that influence the effectiveness of teaching, but this work is only a starting point. There is little firm empirical evidence to support conclusions about the effectiveness of specific approaches to teacher preparation...In general, the evidence supports conclusions about the characteristics it is valuable for teachers to have, but not conclusions about how teacher preparation programs can most effectively develop those characteristics.” The 2019 report by the American Association for the Advancement of Science titled

“A Synthesis of Research on and Measurement of STEM Teacher Preparation” affirms this absence, stating, “Strikingly, our search yielded virtually no published research that explored issues related to the improvement of educator preparation programs” or on the accountability of STEM teacher preparation programs [6]. Further, most of the research in science teacher preparation is on K-8 science teacher preparation, not on secondary science teacher programs [5].

One of the major areas lacking systematic research is the relative quality of different pathways for PTE in the U.S., e.g., whether teacher preparation occurs in an undergraduate program, graduate program, or alternative certification pathway. The NRC report states that “the distinctions among pathways and programs are not clear-cut and there is more variation within the ‘traditional’ and ‘alternative’ categories than there is between these categories.” The report “found no evidence that any one pathway into teaching is the best way to attract and prepare desirable candidates ... This finding does not mean that the characteristics of pathways do not matter; rather, it suggests that research on the sources of the variation in preparation, such as selectivity, timing, and specific components and characteristics, is needed.” The AAAS report states similarly that “The current body of research...does not help us understand how particular parts of programs contribute to valued outcomes given the rest of the program.” [6] Another one of the major areas lacking systematic research is what coursework is needed for teacher education: it is not clear what elements of coursework are important for teacher preparation, the number of courses taken by prospective teachers varies dramatically, and a large number of courses (even as many as required for a major in the discipline) is not necessarily an indicator of teacher quality [5]. Overall, the available research findings “do not cohere to develop the field’s understanding of how programs function and how they relate to a broad range of valued outcomes” [6].

The primary goal of the current paper is to set out the beginnings of, and current opportunities for, systematic research in PTE in the U.S. The PTEPA rubric is well suited to address certain important gaps in research on teacher education programs, including program creation and sustainability, recruitment of physics teacher candidates, and qualities of effective physics teachers. There are other important areas that the PTEPA rubric does not measure: for example, it does not connect features of teacher preparation programs to eventual K-12 student outcomes. Within such limitations, the PTEPA rubric can help measure the extent to which thriving PTE programs have the features that existing reports and professional societies recommend as being important. Below we highlight several features that have been identified as important for PTE programs.

B. PTE programs should secure institutional commitment and strong leadership for robust program creation and maintenance

Multiple studies in physics teacher preparation have found that successful U.S. programs have strong leadership, including at least one person who secures funding and personnel benefiting PTE and negotiates with the institution for changes beneficial to PTE [2,7–9]. Studies in physics suggest that thriving PTE programs need at least one “champion” who is a member of the physics faculty. Such leadership is likely particularly important in disciplinary teacher preparation because of the problem of small numbers: colleges or schools of education are less likely to attend to the specific needs of physics (or other STEM) students seeking to become teachers given their broader priorities in teacher preparation. Champions are more likely to be successful when supported by co-leaders and/or a team who can share the work, provide mutual support, and attack the problem from different angles [10]. PTE program leaders act at a variety of institutional, regional, and national levels with a variety of different constituencies to advance PTE, including direct service to physics teachers, creation and maintenance of PTE programs, departmental leadership and institutional engagement benefiting PTE programs, and national advocacy for PTE. The PTEPA rubric measures many of these features, providing systematic evidence as to whether thriving PTE programs have the features that the literature says they need.

These same studies find that U.S. programs that sustained increases in the number of physics teachers educated per year have substantial institutional commitment to PTE, meaning that the institution supports both the leader(s) and the program. Again, such commitment is likely important given competing priorities for the institution, and that preparing future physics (or STEM) teachers does not typically constitute the regular duties of any unit or person. Support for the leadership team may take a variety of forms, including a mandate to pursue PTE as part of one’s regular teaching, research, and/or service, or establishment of infrastructure supporting PTE (such as a STEM education center or institute). Support for the program may also take a variety of forms, such as salaries for undergraduate early teaching experiences in physics or creation of faculty lines for the specific purpose of supporting PTE.

The PTEPA rubric includes multiple items about specific activities and structures that indicate institutional commitment and strong leadership. In this way, the rubric offers a means to document the extent to which thriving PTE programs are associated with these supposedly necessary features.

C. PTE programs should actively recruit to increase participation

As stated earlier, there is a severe shortage of high school physics teachers across the United States. According to one

major report, a likely factor contributing to these low numbers is that “most physics departments do not carry out substantial formal recruitment efforts specific to their PTE program... Few [departments] directly encouraged students to become high school teachers or sought and identified students who had an interest in teaching” [2]. Another report emphasizes the need to address misconceptions about the teaching profession [11]. Such observations suggest that PTE programs need to conduct active and informative recruitment in order to increase participation, including recruiting undeclared undergraduates and partnering with local school districts [12].

The PTEPA rubric includes multiple elements intended to measure this kind of activity, supporting systematic research to learn what recruitment activities various programs engage in and whether they actually result in increased program participation.

D. PTE programs should offer strong disciplinary preparation to support teacher effectiveness

There is robust agreement that disciplinary preparation is critical for teacher effectiveness, including in physics [5,13–18]. A report by the National Science Teachers Association reviews the extensive literature arguing for strong content preparation for science teachers, while acknowledging that the evidence for specifics is limited (e.g., what depth and breadth of content knowledge teachers need) [19]. Another major report documents solid support for subject matter knowledge for teachers, but admits that it is unclear what constitutes adequate knowledge [13]. The NRC report “Preparing Teachers” reemphasizes the lack of adequate research in this area, stating, “We could find so little detailed information about that preparation, that we cannot answer the question of how well current practice fits the consensus standards” [5].

Pedagogical content knowledge is recognized as a distinct category of disciplinary preparation, including skills to teach using inquiry and a base in scientific practices emphasized by current national science standards [5,20]. These skills are not necessarily gained in upper-division physics courses [5]; instead, some sources assert that teachers should have significant discipline-specific pedagogy experiences, such as multiple physics pedagogy courses and field experiences in secondary classrooms in that discipline [2,14,21–27]. Some empirical studies support the importance of teachers learning content through inquiry [22,28]. Other sources say that the evidence supporting the need for pedagogical content knowledge is limited, and that it is not known whether coursework, fieldwork, or student teaching is an effective means to acquire it [13].

The PTEPA rubric supports documentation of what disciplinary preparation different programs provide, with items identifying both content preparation (e.g., whether teachers graduate with a major in physics) and pedagogical

preparation (e.g., pedagogical coursework and field work accompanied by pedagogical reflection).

E. PTE programs should offer strong mentoring to support application to classroom practice

Another area of consensus is that PTE programs should offer strong mentoring experiences in the context of field experiences and student teaching, to support the application of lessons learned about teaching to actual classroom practice. Multiple studies support the finding that cooperating teachers, who host teacher candidates during field experiences and supervise them during student teaching, should be both good teachers and good mentors to support their mentees' learning [29,30]. Future teachers should observe the kind of teaching that is required of them, and coursework should be deliberately and tightly integrated with field experiences and student teaching [14,31–33]. Other studies emphasize that both mentoring and field experiences should be in the discipline to support continuing application of physics-specific pedagogical knowledge [14,29]. Mentoring strongly influences a student's eventual teaching style: one study even found that in student teaching, the cooperating teacher's mentorship style, not their teaching practices, was the primary factor in determining if the novice teacher would continue to use the teaching practices promoted in their university-based program [30]. Finally, mentoring after a student achieves certification is important for retention in the profession: a large longitudinal study shows that for teachers overall, mentoring during the early years as a working teacher increases the five-year retention rate by about 15% [34].

The PTEPA rubric reflects the literature's emphasis with multiple items about mentoring, including whether there is a mentor with expertise in PTE, whether there is coordination among the PTE mentor, university supervisor, and cooperating teachers, and whether there is mentoring during induction.

F. PTE programs may need other important activities and structures to flourish

There are many other features of U.S. PTE programs that *seem* to be important, based on the fact that successful U.S. PTE programs have these features. For example, a published set of case studies outlines many possibly important features of PTE programs, including the learning assistant experience, undergraduate research, an evidence-based introductory course, active advising, and Teachers in Residence [9]. The report of the Task Force on Teacher Education in Physics [2] draws from literature and examples of programs to say what is important, including a champion, collaboration between physics and education, a sequence of physics pedagogy courses, and early teaching experiences led by the physics department.

However, in many cases empirical studies are not available to validate the importance of these features. Instead, much existing research is in terms of exemplars: descriptive studies of programs that model some desirable aspects of physics teacher preparation. There has so far been little opportunity for comparative studies of different programs to learn which type is more effective for some particular goal, or whether the presence of a particular feature makes a measurable difference to a desired outcome. For example, does an evidence-based introductory course boost the number of eventual teacher graduates? Does student community-building support teacher identity formation? The PTEPA rubric supports measurement of specific features of programs, supporting identification of which activities and structures programs should have and assessment of whether they have them.

III. METHODS: THE PTEPA RUBRIC

The PTEPA rubric was developed by examining literature and instruments to create a draft rubric, and then applying that draft rubric in site visits (both virtual and in-person) to 8 thriving programs. Thriving programs were defined as programs at large U.S. institutions which consistently graduate 5 or more highly qualified future physics teachers each year, and thus were consistent awardees of the “5+ Club” award from PhysTEC. The thriving programs chosen for this study were selected to be diverse: Half had received PhysTEC funding in the past and half had not, and the programs had a wide variety of structures. The list of programs visited is shown below; an asterisk indicates programs which had received PhysTEC funding.

- University of Texas at Austin
- University of Colorado Boulder*
- Brigham Young University
- California Polytechnic University
- San Luis Obispo*
- Georgia State University*
- Rowan University*
- Rutgers University
- Stony Brook University

For each program, one of us (S. V. C. or R. E. S.) conducted a series of interviews over a two-day visit (virtual or in-person), with a variety of program staff which included program leaders, administrators, current and former students, Teachers in Residence, course instructors, partners in physics and education, and so on. The rubric was iterated after each site visit, and again several times once aggregate data were available across programs. Thus, the rubric was developed to reflect an empirical description of what was observed at these programs, rather than to identify *a priori* best practices. More detail about the development process is given in Refs. [1,3,4].

The current version of the rubric consists of 89 items, each with three levels (developing, benchmark, and “exemplary,” plus not present), with benchmark representing a recommended level of achievement. Some of the items are identified as “prevalent,” indicating that the majority of thriving programs achieved at least benchmark level on the item with some degree of certainty [3]. The items are organized into six standards: (1) Institutional Commitment, (2) Leadership and Collaboration, (3) Recruitment, (4) Knowledge and Skills for Teaching Physics, (5) Mentoring, Community, and Professional Support, and (6) Program Assessment. Each standard has three or four components within it that address specific subtopics. The version of the rubric as of this writing is version 2.0 (released August 2018); the latest version can be obtained at the website in Ref. [35].

In addition to ratings on the rubric, rich field notes were taken from each program based on the interviews and other documentation. These notes were coordinated into a body of evidence for the ratings assigned on the rubric for each program. These qualitative results, as well as the quantitative analysis of ratings across programs, are analyzed in this paper to provide insights into the practices of thriving programs. Results reported here reflect the state of each program at the time of the program visit; some program structures and elements have changed since that time. Program visit timing is identified for each program in Ref. [1].

The PTEPA rubric reflects practices and structures of PTE in the U.S. Other countries’ PTE programs share many features with those of U.S. programs, including the importance of training in physics content, pedagogy, and nature of science, field work, and knowledgeable mentors [24,36]; in these areas (most strongly represented in standards 4 and 5), we might expect the PTEPA rubric to align with both U.S. and non-U.S. PTE programs. However, other features of PTE are often distinctively different outside the U.S., and the corresponding parts of the rubric may not align with non-U.S. PTE programs. For example, in-service teachers’ prestige, salaries, and mentoring opportunities vary widely among countries, which would likely affect PTEPA rubric results for recruiting practices (standard 3), early teaching experiences (standard 3), and mentoring and community support (part of standard 5). Additionally, PTE outside the U.S. is often based in colleges or schools of education (or the equivalent) and led by professionals who specialize in physics teacher education [24], whereas in the U.S., PTE is most often led (or co-led) by physics faculty: these differences might produce different rubric results in standards 1 and 2. In a pilot application of the rubric to a single program outside the U.S., that program was rated similarly to thriving U.S. programs in many areas, but lower than thriving U.S. programs in recruitment activities, early teaching experiences, and in-service mentoring, possibly because those

have a different relationship to the PTE program in its social and institutional context.

IV. RESULTS: WHAT DO THRIVING PHYSICS TEACHER EDUCATION PROGRAMS DO?

In what follows, the eight programs studied are mostly referred to with anonymous identifiers (A–H). However, each finding is also illustrated and supported with examples from particular named programs.

A. Thriving programs are strong in multiple areas, and the pattern of strength varies

Thriving programs are strong in multiple standards, suggesting that such programs have a range of high-quality practices and structures rather than excelling only in narrow areas. Figure 1 shows the strength of each program in each standard: for each program, standards in which 50%–75% of the items were rated benchmark or higher are indicated with a light gray box, and standards in which over 75% of items were rated benchmark or higher are indicated with a dark gray box. As shown in the columns of Fig. 1, seven of the eight studied programs (A–G) have a majority of items rated benchmark in at least five standards. Several programs also have a majority of items in multiple standards rated as exemplary (not shown). These ratings are likely positively biased in that the rubric was developed based on the activities and structures at these very programs; future thriving programs might not align so closely with the rubric.

Our holistic observations of specific programs support the finding that thriving programs are strong in multiple standards. For example, the UTeach program at the University of Texas-Austin is strongly supported at the institutional level, including funding from both the College of Natural Sciences and the College of Education to support multiple Master Teachers and staff (standard 1). The PTE program team is led by two faculty co-directors, one in physics and one in education, with several members of the team having substantial positional power (current or former deans) and substantial expertise in physics or physics education research (standard 2). Active recruiting mechanisms include a free one-credit course that introduces students to teaching as a profession by having them design short lessons for elementary students (standard 3). Physics-specific pedagogy is infused into otherwise STEM-broad teacher education courses thanks to program staff with deep expertise in physics and physics education research (standard 4). Multiple Master Teachers provide individual advising, and the program advances supportive community structures both among teacher candidates and with working teachers (standard 5). The program also tracks a wide variety of program metrics, including graduation and persistence data, alumni feedback, and student learning outcomes, and uses these data to inform continuous

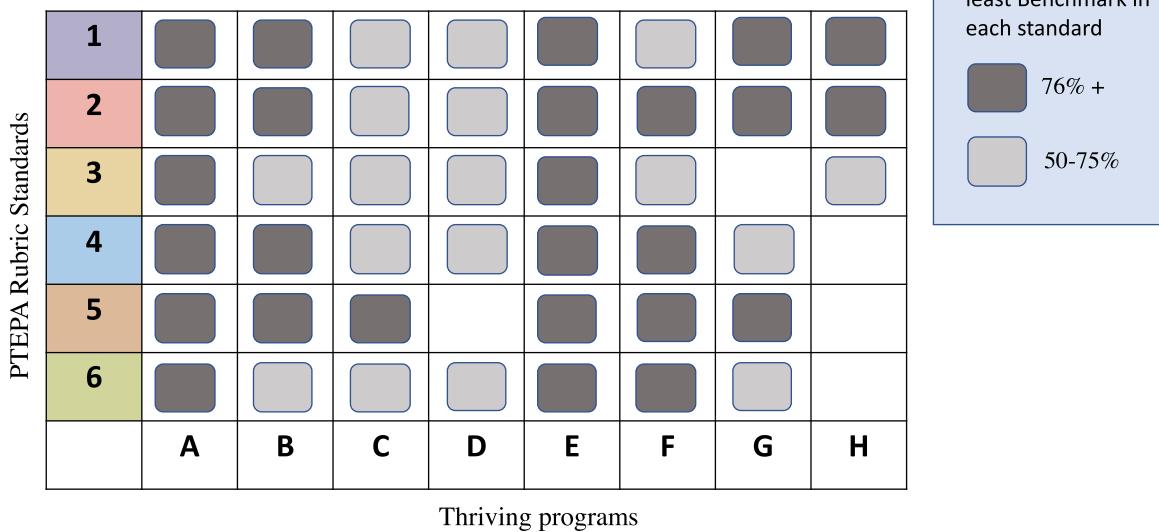


FIG. 1. Percent of ratings at least benchmark level for each standard. Percentages are calculated by adding the percent benchmark and percent exemplary. Those for which 76% or more of items are rated at least benchmark are shaded dark gray; those for which 50%–75% of items are rated at least benchmark are shaded light gray.

improvement (standard 6). Thus, UTeach Austin demonstrates strength across multiple areas of the rubric, as did several other studied programs.

While thriving programs tend to be strong across multiple standards, it is also evident from Fig. 1 that different programs are strong in different standards (no two columns are the same). Additionally, programs D, G, and H have less uniformly positive ratings, showing that there is variety in how thriving programs achieve their strength. This variety among programs is consistent with the literature on teacher education programs, which cited the great variation among programs as a reason for the paucity of empirical research [5].

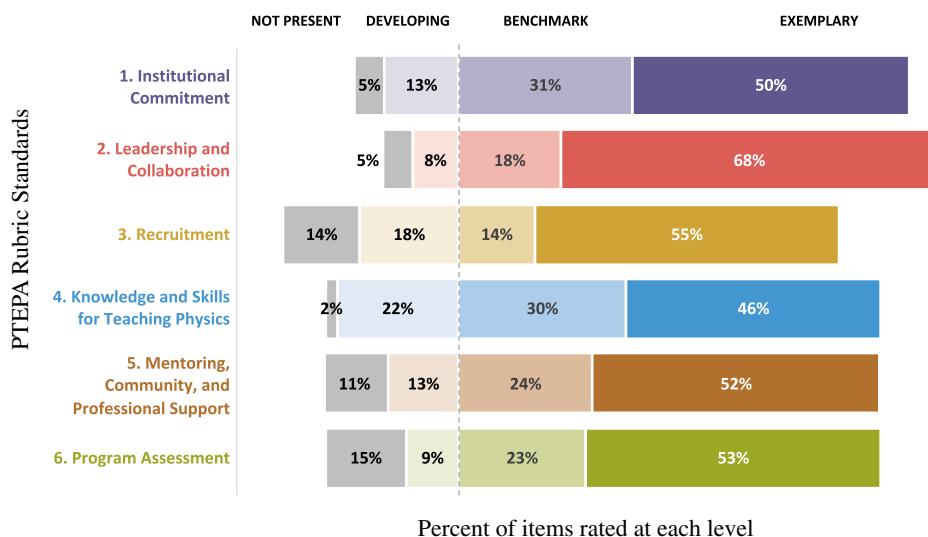
B. The most consistent strength among thriving programs is in institutional commitment, leadership, and collaboration among partners in education and physics

The first two standards on the rubric, Institutional commitment and Leadership and collaboration, are consistently strong among the studied programs. Figure 1 shows that all programs had a majority of items rated at least benchmark in these two standards (those two rows are entirely shaded), and that this is not true of any of the other standards. Figure 2 shows the items on each standard aggregated across all programs studied. Out of all items in standards 1 and 2, 81% were rated at least benchmark in standard 1, and 86% at least benchmark in standard 2. Very few items were at the developing or not present level in these standards. This finding is consistent with existing research on PTE, which has emphasized the leadership

team and institutional commitment as critical features for establishing and sustaining programs [2,7,9].

As supportive evidence for this finding, these rubric data match qualitative observations of specific programs. For example, at the University of Colorado Boulder, PTE is led by a tenured faculty member with extensive expertise in physics education and physics education research, engagement in national PTE networks, and a national reputation for leadership in STEM educational improvements. She is supported by two Master Teachers, who are funded partly by the institution and partly by an endowment. At Rowan University, the PTE team has included the dean of the College of Science and Mathematics, the Physics Department chair, two other faculty in physics, and two Teachers in Residence; multiple team members have decades of high school teaching experience, and one is also a physics education researcher.

Standards 4, 5, and 6 have similar patterns of ratings across all programs studied, with each standard having 76% of items rated at least benchmark (Fig. 2). The least consistently highly rated standard is Recruitment, with only 69% of items rated at least benchmark (Fig. 2), and only two programs have 76% or more of the items rated benchmark (Fig. 1, row 3). The recruitment standard contains a great many items representing a wide variety of recruitment strategies and markers (ranging from number of physics majors, to program reputation, to departmental activity supporting student career exploration, to availability and attractiveness of early teaching experiences, to streamlined and accessible pathways to a teaching career); given this diversity, not all items are likely to be achieved at a high level at even a thriving program.



Percent of items rated at each level

FIG. 2. Percent of items rated at each level across all eight studied programs, by standard. From the left are shown not present (gray), developing (light shaded), benchmark (medium shaded), and exemplary (dark shaded). Percentages are out of the number of available ratings per standard (No. of items/standard \times 8 programs).

Recruitment challenges experienced at many programs include weaknesses in undergraduate physics teaching (which can deter students from considering a teaching career), physics faculty who discourage careers in physics teaching, and the difficulty of attaining licensure within a four-year degree program.

C. Thriving programs are consistently strong in specific areas (rubric components)

Below, we describe some of the specific areas in which thriving PTE programs are observed to excel. This analysis is at the level of PTEPA rubric components. Components are a particularly promising level for analysis in that they seem to be more tightly focused than standards: data from the 8 thriving programs indicate that a majority of components (11 out of 19) show consistent results across a majority of items for a majority of programs, suggesting that they may reflect relatively coherent concepts in PTE [3].

Figure 3 shows the ratings across studied sites, by component. The most consistently strongly rated components are 1C: Resources, and 2B: Program Team Attributes, with over 90% of items rated at least benchmark; this is consistent with the fact that standards 1 and 2 are the most consistently strong standards. The strength of component 6C (Communication to Stakeholders) reflects the indication that a strong institutional commitment (standard 1) is generated and sustained by active and visible communication about the project [10].

The pattern of strong ratings on these components is validated by the body of qualitative evidence from specific programs. For example, the PTE program at Cal Poly San Luis Obispo excels in component 1C (Resources) partly because STEM teacher education is supported by substantial, stable institutional funding for the personnel that run

the programs; there is also significant external funding, dedicated space for the program in a central location, and staff supporting day-to-day operation. Rutgers University's PTE program excels in this same component partly due to its longevity, which creates visibility and a strong reputation that draw students to the program, and has over time produced a large number of local teachers to act as host teachers for student teaching. The program at Georgia State University is strong in component 2B (Program Team Attributes) thanks to a leader with positional power (associate chair) with broad knowledge of physics education, supported by a co-leader in the College of Education and Human Development, as well as a part-time Teacher in Residence. Another example of strength in component 2B is the program at Brigham Young University, led by a physics faculty leader with decades of high school physics teaching experience, in collaboration with a different full-time Teacher in Residence each year who is loaned by a local school district.

Several other components outside of standards 1 and 2 were also strongly rated: we discuss these below, in terms of alignments to the teacher education literature. The lowest-rated component overall is 3D: Streamlined and Accessible Program Options, reflecting the difficulty of satisfying the many requirements for both a physics degree and licensure [37]. The next lowest-rated component is 5C, In-service Mentoring and Support, possibly reflecting the fact that most programs have only scant resources to support program alumni.

D. Thriving programs are consistently strong in specific activities and structures (rubric items)

In addition to addressing areas such as those described above, teacher education programs should have certain

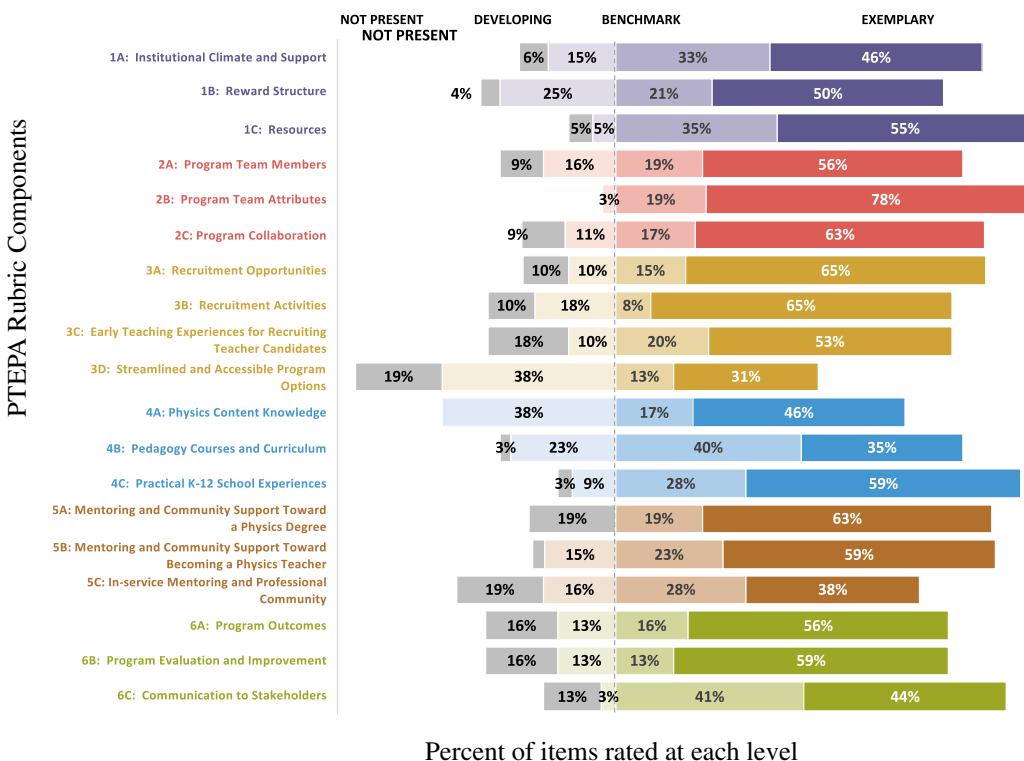


FIG. 3. Percent of items rated at each level across all eight studied programs, by component. From the left are shown not present (gray), developing (light shaded), benchmark (medium shaded), and exemplary (dark shaded). Percentages are out of the number available ratings per component (No. of items/component \times 8 programs).

specific activities and structures. These are represented at the item level of the PTEPA rubric. About half of the items on the rubric are prevalent, meaning that the majority of studied sites achieved benchmark level (or higher) on those items, with a high confidence of the validity of that rating [3]. The prevalent items represent the practices and structures that are most common to thriving programs, and therefore may be the most important to emulate. However, prevalent items are subject to multiple caveats. For example, one of the prevalent items measures future teachers' engagement with a physics pedagogy course. The prevalence of this item may indicate that a physics pedagogy course is common to thriving PTE programs and therefore important to emulate. Alternatively, it may be that the prevalence of this item is sensitive to the wording of the scale points, the specifics of the prevalence selection criteria, or the nature of the programs studied so far. More research will be needed to learn the relationship between prevalent items and effective program structures.

E. Thriving programs are not consistently strong in all specific activities and structures (rubric items)

At the item level, the thriving programs studied are not uniformly exemplary. In fact, all thriving programs had many items rated as not present or developing. As shown in Fig. 4, the eight thriving programs studied so far had up to

18% of items rated as not present and up to 32% of items rated as developing. Among all the studied programs, the percentage of items rated benchmark or higher ranged from 66%–92%; the percentage rated exemplary ranged from 35%–72%. Additionally, as shown in Fig. 2, some thriving programs did not have a majority of items at benchmark level within particular standards. This result reflects the reality that (i) PTE programs vary widely in their activities and structures and (ii) what works is likely context dependent, reflecting local conditions at the institutional and state level.

To characterize the level of variability in rubric ratings across thriving programs, we made use of a measure from ecology, Simpson's diversity index (SDI) [38], which measures the abundance of species (n ; in our case, scale points) in a population (N): $SDI = 1 - \sum(n/N)^2$, where n is the total number of times any of the four scale points is observed in an item, component or standard across all programs studied, and N is the total number of ratings available ($N = 8$ for an individual item, or eight times the number of items in a component or standard). The degree of "diversity" for the rubric is the degree to which ratings on the thriving programs are clustered on a single level (less diverse; lower SDI value) or spread evenly across levels (more diverse; higher SDI values). The SDI for our data ranges from 0 to 0.75, but was normalized to range from 0 to 1.

The average SDI was 0.67 for rubric items, 0.80 for components, and 0.84 for standards, and most SDI values

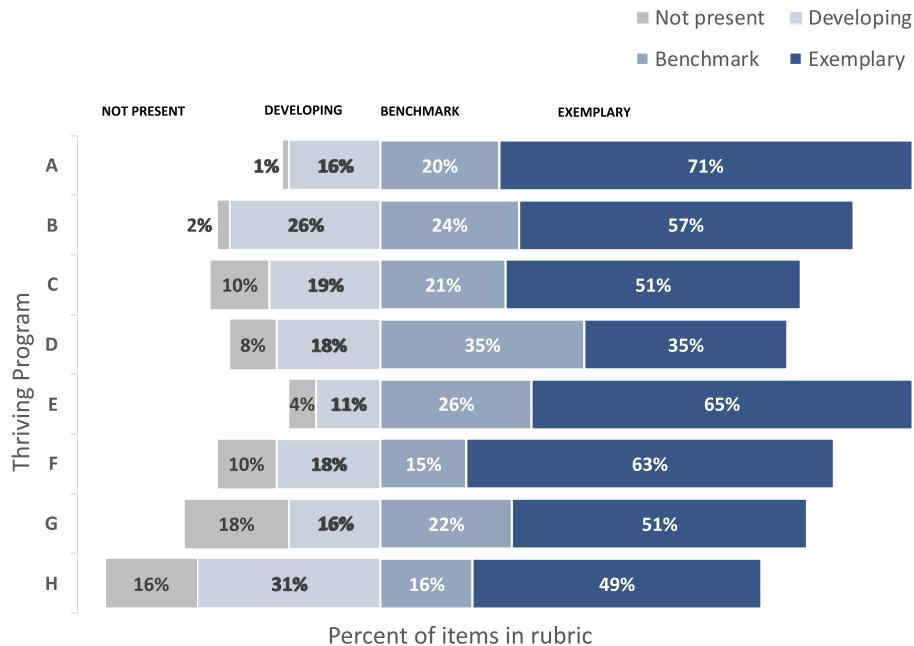


FIG. 4. Percent of items rated at each level across all eight studied programs, on the entire rubric. Percentages are out of the number available ratings per standard (No. of items total \times 8 programs).

for individual items, components, and standards are close to these averages. These results suggest that thriving program ratings are generally spread fairly evenly across the scale points, meaning that the studied programs have diverse strengths on the elements represented by the rubric.

There are, however, a few exceptions of items on which thriving programs were more strongly clustered (SDI values at least two standard deviations lower than the average): All programs received exemplary ratings on the items “Personal Motivation to Improve PTE,” “Professional Engagement in PTE,” “Physics Teaching Ambassador,” and “Physics majors.” The fact that these items are areas of strength for all of the studied programs suggests that they may be necessary features of thriving programs.

V. ALIGNMENT OF THRIVING PROGRAM STRENGTHS AND WEAKNESSES WITH TEACHER EDUCATION LITERATURE

Many of the consistently strongly rated components and items align well with what the literature says that PTE programs should do, indicating that (i) the PTEPA rubric measures activities and structures that are known to be important, and (ii) overall, the studied thriving physics teacher programs have the strengths that the literature says they should have. Below, we revisit the areas of teacher education literature that were described in Sec. II.

A. Institutional commitment and leadership

As stated above, the most consistent strengths among thriving programs are in institutional commitment,

leadership, and collaboration among partners in education and physics. These strengths are evidenced in the consistently high ratings on standards 1 (Institutional commitment) and 2 (Leadership and collaboration). Existing research on PTE, in particular, has emphasized the leadership team and institutional commitment as critical features for establishing and sustaining programs [2,7,9]; thus, this is a major area of alignment between the PTE literature and the findings of this study.

B. Recruitment

The PTEPA rubric includes multiple elements intended to measure recruitment activity at thriving programs, aligned with the PTE literature. For example, component 3B includes items such as whether the program includes a physics teaching ambassador, whether students get accurate information about the career benefits of teaching [11], and whether the physics department exposes students to diverse career options, including physics teaching [39,40]; item 6A-2 documents annual recruitment into the PTE program. These elements are consistently strong at the studied sites, as the teacher education literature suggests that they should be.

Component 3A, Recruitment Opportunities, is also relatively strong across studied programs. This component addresses whether there are structures present that enable significant recruitment, such as a large pool of physics majors or physics-aligned majors and a strong recruitment network. Literature in PTE suggests that among *all* PTE programs, only a small fraction of the variation in the number of PTE program graduates is attributable to program size [2]; our data suggest that among *thriving*

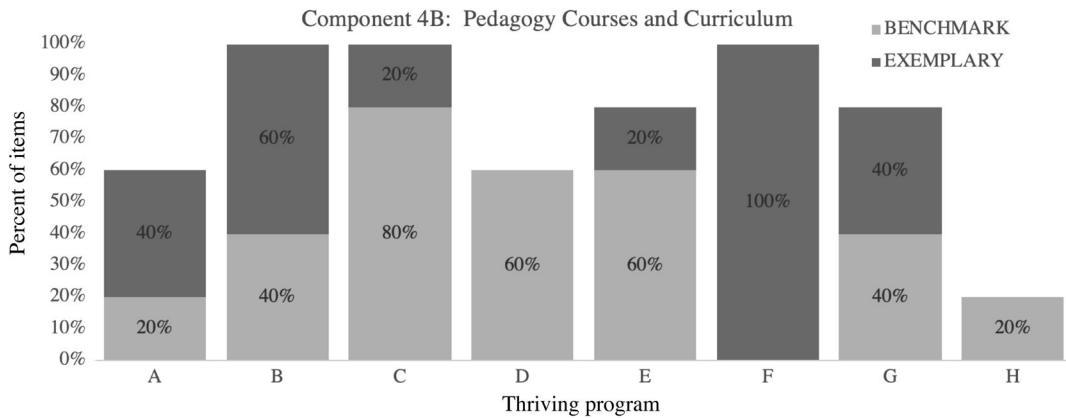


FIG. 5. Percent of items in component 4B rated at benchmark level (light gray) or exemplary level (dark gray) at each of the studied programs (A–H). Percentages are out of the number of items in component 4B, which is five.

programs, a large pool of physics majors may yet be an important piece of institutional context. Having such a recruitment pool and network available is important for generating many future physics teachers, and is distinct from the active and informative recruitment activities that PTE programs need to do to increase participation [12].

C. Disciplinary preparation

The PTEPA rubric documents disciplinary preparation in terms of both content knowledge and pedagogical content knowledge, both of which are major emphases in research on teacher education [5,13,15–18] and research on PTE specifically [14]. Component 4A, Physics Content Knowledge, includes whether a physics minor or equivalent is required for physics teacher candidates (item 4A-1, Physics Degree for Teacher Candidates); this item is prevalent, suggesting alignment with the PTE literature. However, other items in that component are less strong (4A-2, Introductory Physics Course Pedagogy, and 4A-3, Student Research for Teacher Candidates), suggesting either that thriving programs do not reliably ensure that physics teacher candidates have strong physics content knowledge or that those items do not actually represent the elements of content knowledge preparation for those programs.

Pedagogical content knowledge is measured by an array of items in component 4B, Pedagogy Courses and Curriculum, and 4C, Practical K-12 School Experiences. Component 4B has 75% of items rated at least benchmark across the studied programs and 4C has 87% (Fig. 3), suggesting that this study aligns with the PTE research in this area. However, there is a lot of variability among programs: Figure 5 shows that in component 4B, studied programs ranged from 20% benchmark (and no exemplary) to 100% exemplary, with many other combinations in between. This may reflect the diverse means by which

programs pursue pedagogical content knowledge for their teacher candidates: for example, Rutgers University emphasizes disciplinary content in certification coursework (Item 4B-3), while the University of Colorado emphasizes participation in learning assistant programs (Item 4B-5). In practice, we found a wide array of pedagogical preparation structures at the studied sites, and found it particularly challenging to reflect this diversity in a concise set of items. For example, when pedagogy courses are located in the college or school of education, there may either be limited opportunity for physics-specific pedagogy (as at some institutions in this study), or there may be deep integration of physics pedagogy into certification courses (as at Rutgers University).

D. Mentoring

The PTEPA rubric reflects the literature's emphasis on mentoring through multiple elements, including both mentoring toward becoming a physics teacher (component 5B) and mentoring during induction (component 5C). In addition, component 4C (Practical K-12 School Experiences) includes items describing the quality of cooperating teachers and the university supervisor for field experiences. Components 4C and 5B are consistently strong in the studied programs (Fig. 3), suggesting that what these thriving PTE programs do is in alignment with the teacher education literature. Component 5C is less strong across programs (see Fig. 6), reflecting the reality that most PTE programs have limited resources for supporting teachers during induction. In general, more established (e.g., longer-running, institutionally supported) programs have stronger induction practices, suggesting that this kind of activity may be developed by mature programs in response to the needs of growing pools of alumni. However, this activity may be influenced by local conditions, such as whether teacher induction is supported by state agencies (as in New York).

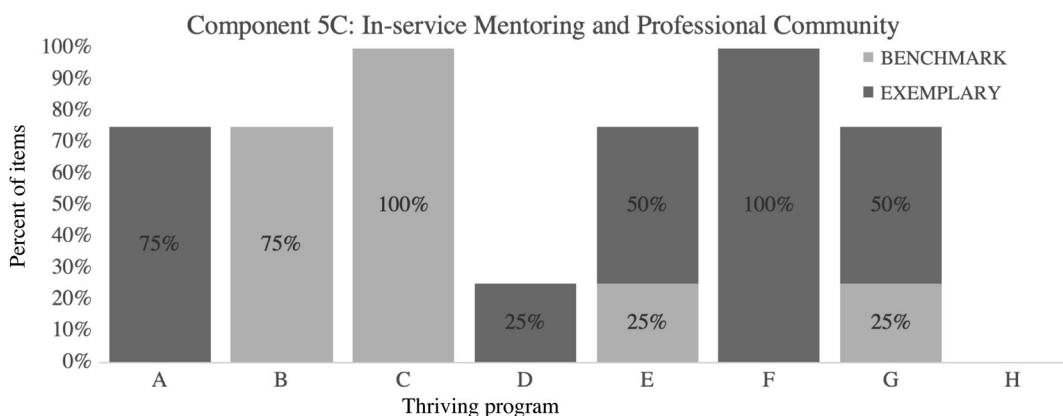


FIG. 6. Percent of items in component 5C rated at benchmark level (light gray) or exemplary level (dark gray) at each of the studied programs (A–H). Percentages are out of the number of items in component 5C, which is four.

VI. SUMMARY, LIMITATIONS, AND FUTURE RESEARCH

The Physics Teacher Education Program Analysis rubric characterizes the practices and structures observed at thriving programs, supporting inquiry into which features of U.S. PTE programs are important for success in which contexts. Initial application of the PTEPA rubric to eight thriving programs demonstrates that (1) thriving programs are strong in multiple areas (standards and components), and the pattern of strength varies across programs; (2) the most consistent strengths among thriving programs are in institutional commitment, leadership, and collaboration among partners in education and physics (standards 1 and 2); (3) thriving programs are strong in some of the specific elements (components) indicated by the PTE literature; (4) thriving programs are consistently strong in certain specific activities and structures (the prevalent items); and (5) thriving programs are not strong in all specific activities and structures (items).

Significant limitations of this study include the fact that there are very few thriving PTE programs (about 27 that we were aware of at the time of the study), meaning that there is a limited basis for generalizations. Another limitation is that only eight of the known thriving programs were studied: we selected for large programs, programs whose teacher graduation rate was steady or increasing, and for a mix of programs that had and had not been funded by PhysTEC. Including different programs in the study, including those at smaller institutions, might have shaped the rubric differently. We also did not control for differing state requirements (such as single subject licensure) which likely has a strong impact on local program structure. All eight programs studied are U.S. institutions; this aligns with the purpose of the study, but limits our findings to PTE in the U.S.

Future research will be greatly enriched by programs that use the PTEPA rubric and share their ratings with PhysTEC, as well as collaboration with a wide range of

research partners with varied interests and skills. Some of this research should be aimed at continuing to improve the PTEPA rubric to better support research (e.g., whether there are missing items, whether the rubric is reliable for self-study, modifying the rubric for valid use at smaller institutions, or whether there ought to be a minimum number of items per component to enhance reliability). Other research has strong potential to guide future PTE program development, such as the following:

What are the most common program features across the U.S.?—Identifying prevalent items with confidence would tell us which features are successfully created and sustained by a variety of programs, which would indicate their feasibility for programs under development. Such research might also reveal whether certain elements of the PTEPA rubric tend to appear together (and therefore are likely related), and whether there are common patterns of PTEPA rubric results (perhaps associated with distinctive types or “profiles” of PTE programs, such as STEM-broad or physics specific, or undergraduate-only versus graduate programs).

What is the influence of context (department, institution, state) on the features of strong programs?—The PTEPA rubric might allow comparisons between program locations (physics department versus colleges or schools of education), models (UTeach versus PhysTEC), types of institution (liberal arts institution versus research university), or broader contexts (states with composite certification versus single-subject licensure). Learning the influence of context would help programs prioritize activities to support the highest possible physics teacher graduation rate and help to better differentiate between different certification pathways. Application of the rubric to non-U.S. programs could also help to elucidate the differences between the U.S. and other countries in terms of PTE.

To what degree are the features we think are necessary for graduating teachers actually necessary?—The most desirable area of PTEPA rubric research would link specific rubric elements (standards, components, or items) to high

teacher graduation rates, potentially challenging accepted wisdom about which features are really needed. Such predictive validity would allow the rubric to provide significant guidance to teacher education programs. For example, what recruitment activities actually result in increased participation? What should large STEM-broad teacher preparation programs do to increase their numbers of physics teacher graduates specifically? Addressing these research questions would help the community place its effort where it will have the most impact.

The PTEPA rubric is at the beginning of its usefulness as a tool for research and systematic improvement of PTE programs. Future research should both increase its validity and expand the associated knowledge base.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grants No. PHY-0808790 and No. PHY-1707990 and the APS 21st Century Campaign. We thank Monica Plisch, Director of Education and Diversity at the American Physical Society, for her leadership and intellectual contributions to the development of the rubric and interpretation of its findings. We thank the gracious hosts and teams at the programs visited for the study for hosting visits, participating in interviews, and reviewing the PTEPA

rubric results and reports: Michael Marder (University of Texas at Austin), Valerie Otero (University of Colorado Boulder), Duane Merrell (Brigham Young University), Chance Hoellwarth (California Polytechnic University, San Luis Obispo), Brian Thoms (Georgia State University), Karen Magee-Sauer (Rowan University), Eugenia Etkina (Rutgers University), and Keith Sheppard (Stony Brook University). Additionally, we acknowledge the contributions of the following: Gay Stewart (West Virginia University), Stamatis Vokos (California Polytechnic University, San Luis Obispo), and Wendy Adams (Colorado School of Mines) provided expert review and feedback on the PTEPA rubric from a PTE perspective. Judy Oakden (Pragmatica) provided review of the instrument from an evaluative rubric perspective. Claudia Fracchiolla (National University of Ireland Galway) created the interactive Excel version of the PTEPA rubric. Justyna P. Zwolak (Joint Center for Quantum Information and Computer Science) provided data analysis and visualization, and Jessica Alzen (University of Colorado Boulder) provided feedback on data interpretation and conducted the SDI analysis. Anthony Ribera (Rose-Hulman Institute of Technology) provided review of academic accreditation processes.

- [1] S. V. Chasteen, R. E. Scherr, and M. Plisch, *A Study of Thriving Physics Teacher Education Programs: Development of the Physics Teacher Education Program Analysis (PTEPA) Rubric* (American Physical Society, College Park, MD, 2018).
- [2] D. E. Meltzer, M. Plisch, and S. Vokos, *Transforming the Preparation of Physics Teachers: A Call to Action. A Report by the Task Force on Teacher Education in Physics (T-TEP)* (American Physical Society, College Park, MD, 2012).
- [3] S. V. Chasteen, R. E. Scherr, and M. Plisch, preceding paper, The PTEPA Rubric: Measuring features of thriving physics teacher education programs, *Phys. Rev. Phys. Educ. Res.* **16**, 010115 (2020).
- [4] R. E. Scherr and S. V. Chasteen, R. E. Scherr, and S. V. Chasteen, Development and validation of the Physics Teacher Education Program Analysis (PTEPA) Rubric, in *Proceedings of the 2018 Physics Education Research Conference, Washington, DC* (AIP, New York, 2018).
- [5] National Research Council, *Preparing Teachers: Building Evidence for Sound Policy* (The National Academies Press, Washington, DC, 2010).
- [6] C. Bell, D. Gitomer, and A. H. McKenna, *A Synthesis of Research on and Measurement of STEM Teacher Preparation* (American Association for the Advancement of Science, Washington, DC, 2019).
- [7] R. E. Scherr, M. Plisch, and R. M. Goertzen, Sustaining physics teacher education coalition programs in physics teacher education, *Phys. Rev. Phys. Educ. Res.* **13**, 010111 (2017).
- [8] R. E. Scherr, M. Plisch, and R. M. Goertzen, *Sustaining Programs in Physics Teacher Education: A Study of PhysTEC Supported Sites* (American Physical Society, College Park, MD, 2014).
- [9] *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices*, edited by C. Sandifer and E. Brewe (American Physical Society, College Park, MD, 2015).
- [10] A. Kezar, *Understanding and Facilitating Organizational Change in Higher Education in the 21st Century*, 1st ed. (Jossey-Bass, San Francisco, 2001).
- [11] Get the Facts Out, Get the Facts Out. [Online]. Available: <https://getthefactsout.org/>. [Accessed: 14-Aug-2019].
- [12] B. C. Clewell and L. B. Forcier, Increasing the number of mathematics and science teachers: A review of teacher recruitment programs, *Teach. Change* **8**, 331 (2001).
- [13] M. Allen, *Eight Questions on Teacher Preparation: What Does the Research Say? A Summary of the Findings* (Education Commission of the States, Denver, CO, 2003).
- [14] E. Etkina, B. Gregoric, and S. Vokos, Organizing physics teacher professional education around productive habit

development: A way to meet reform challenges, *Phys. Rev. Phys. Educ. Res.* **13**, 010107 (2017).

[15] National Research Council, *How People Learn: Brain, Mind, Experience, and School: Expanded Edition* (National Research Council, Washington, DC, 1999).

[16] H. Hill and D. L. Ball, The curious—and crucial—case of mathematical knowledge for teaching, *Phi Delta Kappan* **91**, 68 (2009).

[17] L. Darling-Hammond, *Powerful Teacher Education: Lessons from Exemplary Programs*, 1st ed. (Jossey-Bass, San Francisco, CA, 2013).

[18] L. Darling-Hammond and J. Bransford, *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do* (Jossey-Bass, Wiley, New York, 2005).

[19] NSTA Standards for Science Teacher Preparation. [Online]. Available: <https://www.nsta.org/preservice/>. [Accessed: 14-Aug-2019].

[20] Next Generation Science Standards. [Online]. Available: <https://www.nextgenscience.org/>. [Accessed: 14-Aug-2019].

[21] E. Etkina, Pedagogical content knowledge and preparation of high school physics teachers, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020110 (2010).

[22] L. C. McDermott, P. S. Shaffer, P. R. L. Heron, M. R. Stetzer, and D. L. Messina, Preparing teachers to teach physics and physical science effectively through a process of inquiry, in *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices*, edited by C. Sandifer and E. Brewe (American Physical Society, College Park, MD, 2015), pp. 165–186.

[23] L. C. McDermott, A perspective on teacher preparation in physics and other sciences: The need for special science courses for teachers, *Am. J. Phys.* **58**, 734 (1990).

[24] D. E. Meltzer, Research on the education of physics teachers, in *Teacher Education in Physics: Research, Curriculum, and Practice*, edited by D. E. Meltzer and P. S. Shaffer (American Physical Society, College Park, MD, 2011), pp. 3–14.

[25] R. M. Sperandeo-Mineo, C. Fazio, and G. Tarantino, Pedagogical content knowledge development and preservice physics teacher education: A case study, *Res. Sci. Educ.* **36**, 235 (2006).

[26] J. R. Thompson, W. M. Christensen, and M. C. Wittmann, Preparing future teachers to anticipate student difficulties in physics in a graduate-level course in physics, pedagogy, and education research, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010108 (2011).

[27] A. Berry, P. Friedrichsen, and J. Loughran, *Re-examining Pedagogical Content Knowledge in Science Education* (Routledge, London, 2015).

[28] J. A. Marshall and J. T. Dorward, Inquiry experiences as a lecture supplement for preservice elementary teachers and general education students, *Am. J. Phys.* **68**, S27 (2000).

[29] D. J. Boyd, P. L. Grossman, H. Lankford, S. Loeb, and J. Wyckoff, Teacher preparation and student achievement, *Educ. Eval. Policy Anal.* **31**, 416 (2009).

[30] M. Cochran-Smith, A. M. Villegas, L. Abrams, L. Chavez-Moreno, T. Mills, and R. Stern, Critiquing teacher preparation research: An overview of the field, part II, *J. Teach. Educ.* **66**, 109 (2015).

[31] L. Darling-Hammond, Constructing 21st-century teacher education, *J. Teach. Educ.* **57**, 300 (2006).

[32] L. Darling-Hammond, R. C. Wei, A. Andree, N. Richardson, and S. Orphanos, *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad* (National Staff Development Council, Oxford, Ohio, 2009).

[33] R. N. Steinberg, *An Inquiry into Science Education, Where the Rubber Meets the Road* (Sense Publishers, Leiden, Netherlands, 2011).

[34] Public School Teacher Attrition and Mobility in the First Five Years: Results From the First Through Fifth Waves of the 2007–08 Beginning Teacher Longitudinal Study, 30-Apr-2015. [Online]. Available: <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015337>. [Accessed: 14-Aug-2019].

[35] <http://physstec.org/thriving>.

[36] S. K. Abell, International perspectives on science teacher education, in *Science Teacher Education: An International Perspective*, edited by S. K. Abell (Springer Netherlands, Dordrecht, 2000), pp. 3–6.

[37] S. V. Chasteen, R. E. Scherr, and M. Plisch, What are some challenges observed at thriving programs?, in *A Study of Thriving Physics Teacher Education Programs: Development of the Physics Teacher Education Program Analysis (PTEPA) Rubric* (American Physical Society, College Park, MD, 2018).

[38] E. H. Simpson, Measurement of diversity, *Nature (London)* **163**, 688 (1949).

[39] P. R. L. Heron and L. McNeil, *Phys21: Preparing Physics Students for 21st Century Careers* (American Physical Society, College Park, MD, 2016).

[40] AIP Career Pathways Project, Society of Physics Students, 03-Dec-2014. [Online]. Available: <https://www.spsnational.org/career-resources/career-pathways>. [Accessed: 14-Aug-2019].

[41] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.16.010116>, which includes the version of the rubric referenced in this publication, interview protocols for development of the rubric, a list of items on the PTEPA rubric which do not appear on the TEPA rubric, a list of items removed from the PTEPA rubric, and ratings of all PTEPA rubric items for the eight thriving programs in the study.