

Development and validation of the Physics Teacher Education Program Analysis (PTEPA) Rubric

Rachel E. Scherr¹ and Stephanie V. Chasteen²

¹*Scherr & Associates, 3045 NE 86th Street, Seattle, WA, 98115*

²*Chasteen Educational Consulting, 247 Regal St., Louisville, CO 80027*

The Physics Teacher Education Coalition (PhysTEC) conducted a study of eight diverse thriving physics teacher education programs, defined as programs at large universities that typically graduate five or more physics teachers in a year. The study identified common characteristics of such model programs to support other institutions in emulating them. These characteristics are embodied in a new tool, the Physics Teacher Education Program Analysis (PTEPA) Rubric. This paper documents the development and validation of the PTEPA Rubric as a tool for self-analysis and research. We also document the approach to serve as a potential guide for others who wish to develop evaluative rubrics for complex projects.

I. INTRODUCTION

The Physics Teacher Education Program Analysis (PTEPA) Rubric was developed as a tool to characterize the practices and features of “thriving” physics teacher education (PTE) programs, defined as PTE programs at large universities that frequently graduate five or more highly qualified physics teachers in a year. The rubric thus serves as a taxonomy of PTE program elements, and also describes different levels of achievement to enable programs to measure the relative strength of each element. The rubric is intended to offer programs a catalog or “roadmap” of best practices, to guide programs in self-reflection toward improvement, to provide feedback on their progress, and to provide a means to characterize program growth. It is also a tool to enable future research studies, allowing comparison of different programs, aggregation of information on activities across programs, and investigation of how program elements correlate with physics teacher graduation rates.

This paper describes the development and validation of the PTEPA Rubric, both for the purposes of the PTE community, and as documentation for others developing analytic rubrics. The latest version of the rubric, along with a detailed report describing its use and interpretation, is at <http://phystec.org/thriving> [1].

II. DEVELOPMENT OF THE PTEPA RUBRIC

Using a rubric assumes that a complex entity (such as a PTE program) may be represented in terms of a number of discrete elements (in this case, the elements of the PTEPA Rubric) and that the quality of those elements can be approximately measured through rating on a subjective scale. Creating a rubric is also an opportunity to surface implicit values and desired outcomes in a project, bringing attention to key aspects and creating shared and explicit understanding about goals and achievement [2,3]. Evaluative rubrics have been demonstrated to be valuable for answering such questions of quality even in similarly complex situations [2-4]. Evaluative rubrics have been employed in several other projects, such as the Teacher Education Program Assessment

(TEPA) [5], the PULSE rubrics [6], and several program accreditation and review processes [1].

A. Rubric structure

The PTEPA Rubric focuses on physics teacher education, explicitly avoiding areas which lie within the domain of the school of education (except in areas such as student teaching, where physics-specific support is the focus). Early drafts of the PTEPA Rubric were informed by existing statements of common features of thriving physics (or science) teacher education programs. The first draft was based on the TEPA [5], and the PTEPA Rubric retains some overarching categories from the TEPA. Successive versions of the PTEPA Rubric were reconciled with the PhysTEC Key Components, the report of the Task Force on Teacher Education in Physics (T-TEP) [8], and several reports on strong physics programs and career preparation [1]. However, the most significant effort in developing the PTEPA Rubric went toward shaping the rubric to represent what researchers observed at the eight diverse thriving programs in the study (see below). Development of the structure of the rubric was a significant effort, with items, components, and standards iterated over many versions to find a meaningful organization.

The PTEPA Rubric was eventually structured in a format influenced by the Council for the Accreditation of Educator Preparation (CAEP) Evaluation Rubric [7], using the language and structure of “standards” and “components,” narrative descriptions of those standards and components, and examples of attributes “below,” “meeting,” and “above” a sufficient level for each item.

B. Items and scale points

The PTEPA Rubric development required identifying (1) the dimensions or criteria of performance to be measured, called “items,” and (2) the levels of achievement on those items, called “scale points” [1,4,9]. The PTEPA Rubric represents what is called an “analytic rubric,” where the levels of achievement are described quite specifically for each level of each item [3]. This is in contrast to a “holistic”

or generic rubric in which a single generic scale is used across all items to describe high or low performance. The choice of an analytic rubric is intended to increase the reliability of ratings; however, as has been acknowledged by other authors, it also adds a layer of complexity and challenge in rubric development [2,3,9]. The PTEPA Rubric uses three scale points per item. Originally five levels of achievement were developed, but it was exceedingly difficult to identify and agree upon five distinct levels which progressed linearly and meaningfully distinguished programs, and it was also difficult to assign programs to such fine-grained levels. The final decision of three levels (with the middle level being “sufficient”) is more typical; this was found to be more usable and is in alignment with other developers [10], who indicate that it is harder for a reader to make sense out of more than three levels.

The PTEPA Rubric retains many individual items that originated with the TEPA [5]; other rubric items were added from other instruments, suggested by consultation with experts, and indicated by program observations. Researchers attempted to reconcile features that were suggested by these different sources with items on the rubric. When features from various sources reinforced one another, researchers prioritized including such an item on the rubric; when features did not correspond (e.g., a particular type of course was named as important at one program, but experts indicated that this situation was an anomaly), researchers considered whether that feature might not be included on the rubric. See [1] for a detailed history of many rubric items.

For each of the PTEPA Rubric items, researchers identified concrete hallmarks of the levels at which a program may exemplify that feature:

- *Developing*: The program performs better than a typical U.S. institution of higher education on that item.
- *Benchmark*: The program performs at a recommended level on that item.
- *Exemplary*: The program is among the best-performing on that item.

The “Developing” level was constructed to be achievable by many physics teacher education programs; “Benchmark” is designed to represent satisfactory achievement, and the “Exemplary” level is meant to represent an ambitious yet feasible goal for the item. “Not Present” indicates the program does not meet a minimum level on the item. The purpose of the scale points is to illustrate different levels of achievement for each item in specific, well-defined terms and to set expectations for what constitutes strength in each item. Where possible, scale points are both unidimensional and objective, and levels progress somewhat linearly.

C. Standards and components

The PTEPA Rubric is 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, and each component has two to eight individual items. The ordering of the standards represents a progression from the foundations of a physics teacher education program (institutional setting and program leadership) through a prospective teacher’s experiences (recruitment, student learning, and mentoring), and finally reflection on the program (assessment). Standards and components in the instrument typically represent broader program objectives, whereas items represent more concrete programmatic structures, policies, or actions. Where possible, the items represent directly measurable entities, and achievement of components or standards may be indirectly inferred by the degree of achievement of the items grouped under them.

To avoid repetition across standards, we often disaggregated separate aspects of items so they could appear in the most appropriate standard. For example, Learning Assistant programs have multiple aspects; we placed “Teaching/Learning Assistant participation” within Standard 4: Knowledge and Skills for Teaching Physics (measuring the number of future physics teachers who actually benefit from the experience for improving their teaching skills), but we placed “Availability of early teaching experiences” in Standard 3: Recruitment (acknowledging that how broadly available these experiences are is a measure of their ability to draw students into teaching).

III. THRIVING PROGRAM VISITS

Thriving programs studied were large programs that were consistently members of “The 5+ Club,” an honor that PhysTEC awards to physics teacher education programs that graduate five or more teachers in any given year. The eight institutions studied were selected to represent the diversity of thriving physics teacher education programs. Half of the studied programs had received PhysTEC funding and half had not; including non-PhysTEC sites helped ensure validity of the rubric for all thriving programs by (1) avoiding circular logic (since PhysTEC sites are required to address the Key Components in their programs, including only PhysTEC sites might have led to the rubric resembling the Key Components) and (2) broadening the item pool (including non-PhysTEC sites helped to identify items that were not part of the PhysTEC model, such as program reputation, microteaching, or elements of student teaching). Additionally, the programs had a wide variety of structures (e.g., undergraduate programs, post-baccalaureate programs, UTeach replication programs, programs run by a single faculty leader, programs run by a large multidisciplinary team, and so on). The physics teacher education programs represented in this study are (PhysTEC sites are indicated with an asterisk): 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, and Stony Brook University.

The program visits served multiple purposes, including instrument development, initial validation, collection of data supporting research findings, documentation of strongest elements and common challenges, and case study development [1].

For each program, the researcher completed the PTEPA Rubric by conducting a series of interviews (ranging from 8-19) with program personnel. For a given program, the series of interviews might have included faculty leaders, program staff, physics teacher candidates, Teachers in Residence, administrators, cooperating teachers, program graduates, and so on. Program visits were conducted either through intensive two-day in-person program visits (three programs) or two-day “remote program visits,” consisting of a series of concentrated interviews by telephone and videoconference (five programs). PhysTEC-funded programs were prioritized for all in-person program visits to better justify the use of PhysTEC program resources. In either case, interviews were guided by extensive protocols [1]. Researchers documented the interviews with extensive notes pertaining to each rubric item and component, with special attention to the evidence justifying the rating on each item and triangulation of that evidence among multiple interviewees.

IV. VALIDATION OF THE PTEPA RUBRIC

Like all validation, the validation of the PTEPA Rubric is hypothesis-specific. In this study, validation efforts investigate the hypothesis that the PTEPA Rubric measures features that thriving programs tend to have. As an analogy, we might liken the PTEPA Rubric to a “Health Rubric” that elucidates healthy lifestyle behaviors such as nutrition and exercise, and we might compare a high teacher graduation rate to a healthy weight; in this analogy, the hypothesis would be that the “Health Rubric” measures behaviors that people at a healthy weight tend to have. Additional potential hypotheses regarding the PTEPA Rubric are listed in [1].

A. Alignment with previous work

Alignment with relevant instruments and research findings supports several types of validity of the instrument: *Substantive validity* describes the quality of the theoretical basis for the features thought to be important for thriving physics teacher education programs, and *content validity* describes the extent to which the instrument has all the features thought to be important for thriving programs. The PTEPA Rubric is based on and aligned with other validated instruments, and thus it is likely to inherit the substantive and content validity of those instruments. In addition, the PTEPA Rubric has been extensively reconciled with relevant literature and reports along with researchers’ observations of, and program leaders’ knowledge of, thriving PTE programs.

For example, the TEPA [5], which formed the initial basis for the PTEPA Rubric, was developed through extensive expert feedback and program visits. The TEPA

includes many standards and components similar to those in the PTEPA Rubric, including “Leadership,” “Recruitment,” “Content and Pedagogy,” “Beginning Teacher Support,” and “School Development.” The PTEPA Rubric is also well aligned with the T-TEP report [8] and the associated policy statement, which highlight the areas of collaboration, pedagogy, recruitment, and beginning teacher support as important for physics teacher preparation. The PTEPA Rubric components also correspond well with the PhysTEC Key Components. It is worth noting that the PTEPA Rubric was not explicitly developed to directly align with any of the other reports or instruments referenced in [1] except the TEPA and the Key Components, and it was significantly restructured from the template suggested by the TEPA. The fact that the PTEPA Rubric essentially “rediscovered” many elements recommended in these other sources provides some validation for the content and structure of the instrument.

The areas in which the PTEPA Rubric does not align with other instruments are often specific to *physics* teacher education programs. The areas in the TEPA that do not appear in the PTEPA Rubric are almost entirely within the domain of schools of education or teacher certification. Similarly, since the PTEPA Rubric is focused primarily on supporting increased *numbers* of graduates from PTE programs, and elements of early teacher support are more minimally included (within Standard 5: Mentoring, Community, and Professional Support) even though such aspects are important for career development and retention.

B. Identification of prevalent items

Further evidence of the content validity of the PTEPA Rubric derives from identification of individual items in which many thriving programs are strong. These items are denoted as “Prevalent” items. Prevalent items are those for which (1) at least six of the eight studied programs were rated at least at Benchmark level, and (2) at least one of the following confidence measures is met (to strengthen reliability):

- The item is inherently reliable because it measures an objective quantity, such as the number of faculty leaders or Teachers in Residence.
- At least six of the eight studied programs were rated Exemplary on that item.
- Both members of two pairs of very different types of programs were rated at least Benchmark on that item, indicating that the item is important across contexts (e.g., at both a large STEM teacher education program run from outside the physics department and a small program led by a single physics faculty member).

Prevalent items might indicate these activities are more necessary for a thriving program or might indicate the items are relatively easy for programs to achieve. Some items have low ratings at the studied programs, suggesting that even thriving programs have less success in these areas. This may indicate that these items are less essential to a thriving

program or that they are very difficult to achieve. For full results on PTEPA Rubric items, see [1].

C. Expert Review

Face validity of the PTEPA Rubric describes the extent to which the PTEPA Rubric appears to experts to measure the features that thriving physics teacher education programs tend to have. The PTEPA has been scrutinized by national experts in physics teacher education, including the PhysTEC leadership team, members of T-TEP, the leaders of the thriving programs studied in this report, and researchers who study physics teacher education [1]. These experts raised important questions that influenced revision of the PTEPA Rubric at many stages, from the scope of the instrument to addition and clarification of individual rubric items.

V. SUMMARY

This study has identified elements of thriving physics teacher education programs through an extensive development and validation process that resulted in the PTEPA Rubric. Development activities have included theoretical development, adaptation of related instruments, and studies of thriving PTE programs. Sources of validation have included researcher observations, program leaders' knowledge, literature and reports of thriving physics teacher education programs, and expert review. This systematic development and validation has produced an instrument that reflects the practices and priorities of diverse thriving physics teacher education programs.

We encourage others to consider rubrics as a tool for research and evaluation of complex projects. The explicit identification of dimensions of PTE programs has been valuable for PhysTEC and PTE program leaders. However, development of an analytical rubric for a complex project, with full descriptions for each scale point, is a substantial task. The process of iterating the instrument through user feedback and data from actual programs is time-consuming (the PTEPA Rubric was developed during the course of 18 months of focused effort) but invaluable in creating a usable tool which reflects the values of the community. Such a

development process also conforms to ethical standards by protecting the welfare of stakeholders, by including a range of relevant perspectives, and reducing potential risk from the use of such an evaluative tool (see [1] for more on ethics.)

The PTEPA Rubric is a valuable tool for self-study among PTE programs; program leaders should consider completing the rubric for their own program and using the results for continuous improvement, possibly aligning their activities with those of thriving programs. Caution must be taken in interpreting rubric ratings, however; elements of the PTEPA Rubric are neither independent nor of equal weight. Thus, one should not turn PTEPA Rubric results into numerical "scores" for comparing programs. One also cannot say that a program that has 60% of its items rated as Exemplary is definitively better than one that shows 50% rated as Exemplary or that two programs that have equal PTEPA Rubric results are equally good. Many resources for using the rubric are at the URL above, and in [1].

The PTEPA Rubric is at the beginning of its development as a tool for feedback, research, and systematic improvement of physics teacher education programs. A revised version of the instrument based on continuing validation work has already been released at <http://phystec.org/thriving>. Future research efforts should focus on increasing its validity and expanding the associated knowledge base. We encourage the physics education research community to make use of the rubric as a research tool and to partner with PhysTEC in enhancing its reliability.

ACKNOWLEDGEMENTS

We acknowledge Monica Plisch, Director of Education and Diversity at the American Physical Society, for envisioning this project and her engaged and thoughtful collaboration, as well as many other critical collaborators at the APS. We thank the gracious hosts and teams at the programs visited for the study, G. Stewart, S. Vokos, and W. Adams for feedback on the rubric, and many others; see full report [1]. This work was supported in part by the National Science Foundation under grant numbers PHY-0808790 and PHY-1707990 and the APS 21st Century Campaign.

-
- [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] J. King et al, *J. Multidisciplinary Eval.*, 9, 21 (2013)
 - [3] D. Allen and K. Tanner, *CBE-Life Sci. Educ.*, 5, 3 (2006)
 - [4] E. J. Davidson, *Evaluation methodology basics* (Sage, California, 2005)
 - [5] C. R. Coble et al., *Teacher Education Program Assessment* (2012). http://www.aplu.org/projects-and-initiatives/stem-education/SMTI_Library/TEPA/file
 - [6] K. M. Aguirre et al., *CBE-Life Sci. Educ.*, 12, 4 (2013)
 - [7] Council for the Accreditation of Educator Preparation, *CAEP Accreditation Handbook* (CAEP, Washington DC, 2016)
 - [8] D. E. Meltzer, M. Plisch, and S. Vokos, *Transforming the preparation of physics teachers: A call to action* (American Physical Society, College Park, MD, 2012)
 - [9] J. Oakden, *Understanding the components of evaluative rubrics and how to combine them* (Pragmatica, Wellington, NZ, 2018)
 - [10] M. J. Bresciani, J. A. Anderson, and C. L. Zelna, *Assessing student learning and development* (NASPA, Washington, DC, 2004)