



Teacher candidate performance assessments: Local scoring and implications for teacher preparation program improvement



Kevin C. Bastian^{a,*}, Gary T. Henry^b, Yi Pan^c, Diana Lys^d

^a University of North Carolina at Chapel Hill, Education Policy Initiative at Carolina, Abernathy Hall, Campus Box 3279, Chapel Hill, NC 27599, USA

^b Vanderbilt University, Peabody College, PMB #414, 230 Appleton Place, Nashville, TN 37203, USA

^c Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill, 105 Smith Level Road, Chapel Hill, NC 27516, USA

^d 101 Peabody Hall, Campus Box 3500, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

HIGHLIGHTS

- Locally-scored performance assessments partially-aligned with construct framework.
- Locally-scored performance assessments systematically higher than official scores.
- Locally-scored performance assessments significantly predict teacher outcomes.
- Candidate performance assessments may inform evidence-based program improvement.

ARTICLE INFO

Article history:

Received 5 October 2015

Received in revised form

6 May 2016

Accepted 16 May 2016

Keywords:

Performance assessments
Evidence-based program improvement
Construct validity
Reliability
Predictive validity

ABSTRACT

Locally-scored teacher candidate performance assessments offer teacher preparation programs (TPPs) formative performance data, common language and expectations, and information to guide program improvements. To best use these data, TPPs need to understand the validity and reliability of local scoring and assess whether scores predict candidates' performance as teachers. Examining locally-scored performance assessments, we find that local scores are significantly higher than official scores. However, local scores identify three factors partially-aligned with the assessment's construct blueprint and significantly predict teachers' performance outcomes. These analyses provide a framework for research and highlight the utility of locally-scored performance assessments for evidence-based TPP improvement.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, public concern for the quality of teachers and teacher education has pushed policymakers and accreditation agencies in the United States to hold teacher preparation programs (TPPs) accountable for the effectiveness of their graduates (Crowe, 2011). For example, shortly after the implementation of the federal No Child Left Behind act in 2002, states such as Louisiana, North Carolina, and Tennessee initiated efforts to link teachers' value-

added scores to the TPP from which they graduated (Bastian, Patterson, & Yi, 2015; Gansle, Noell, & Burns, 2012; Henry et al., 2011; Henry, Thompson, Fortner, Zulli, & Kershaw, 2010; Noell & Burns, 2006; Noell, Porter, Patt, & Dahir, 2008; TSBE, 2012, 2013). In 2009, the Race to the Top grant competition mandated that states seeking federal funds commit to publicly reporting TPP's effectiveness on value-added measures and closing low performing TPPs (Crowe, 2011; Henry, Kershaw, Zulli, & Smith, 2012). More recently, the United States Department of Education has proposed regulations that would require TPPs to report a variety of performance measures, including the learning outcomes for graduates' K-12 students (Federal Register 2014-28218, 2014). Likewise, the Council for the Accreditation of Educator Preparation (CAEP), the national accrediting body for educator preparation programs, requires TPPs to demonstrate the impact of their graduates on student learning,

* Corresponding author. Department of Public Policy, Education Policy Initiative at Carolina, University of North Carolina at Chapel Hill, Abernathy Hall, Campus Box 3279, Chapel Hill, NC 27599, USA.

E-mail addresses: kbastian@email.unc.edu (K.C. Bastian), gary.henry@vanderbilt.edu (G.T. Henry), yi.pan@unc.edu (Y. Pan), lys@unc.edu (D. Lys).

classroom instruction, and employer satisfaction (CAEP 2013).

In response to these policies and the desire of teacher educators to prepare more effective beginning teachers, TPPs have begun to reform their preparation practices and engage in continuous improvement efforts. Given the current policy context and its focus on the achievement scores of students taught by TPP graduates, it appears that the success of these reforms will be judged, at least in part, on the value-added scores of TPP graduates. By themselves, however, teacher value-added scores are insufficient to guide TPP reforms for two reasons. First, value-added scores come too late to guide TPP improvement efforts—there are often several years between teacher candidates' preparation and their entry into the workforce and impact on student learning. Second, while measuring one aspect of teachers' effectiveness, the value-added scores of program graduates do not provide information about specific teaching practices that would allow TPP faculty and staff to identify programmatic strengths and weaknesses. While some states and school districts use multiple measures—value-added, classroom observations, evaluation ratings—to assess teacher performance, these still suffer from the first problem—they come too late to guide TPP improvement.

To best drive program improvement efforts, TPPs need data on the performance of their candidates that is timely, identifies multiple domains of teaching effectiveness, and significantly predicts outcomes for teachers-of-record. At least one study suggests that many traditional measures of teacher candidate performance, such as grade point average, licensure exam scores, dispositional ratings, and student teaching ratings, do not meet all these criteria and thus, may be of limited use for evidence-based program improvement (Henry et al., 2013). In recent years, however, many teacher educators have supported the creation and widespread adoption of teacher candidate performance assessments—one of which has received widespread attention, the Teacher Performance Assessment (TPA). The TPA is a portfolio completed by teaching candidates during their student teaching experience that uses video clips of instruction, lesson plans, student work samples, and candidates' reflective commentaries to examine candidates' ability to effectively plan for instruction, teach in their content area, and assess both students and their own teaching. These assessments are scored using rubrics that have been field tested for reliability (Stanford Center for Assessment, Learning, and Equity (SCALE, 2013)).¹

While teacher candidate performance assessments could be used as a high-stakes measure for certification and/or program completion decisions (Duckor, Castellano, Tellez, Wihardini, & Wilson, 2014), performance assessments that are locally-scored by TPP faculty and staff may inform evidence-based program improvement efforts. As argued by Peck and colleagues, locally-scored performance assessments provide TPP faculty and staff with: (1) a common language for discussing candidates' performance; (2) common expectations for teacher candidate performance; (3) a forum for accepting collective responsibility for teacher candidate performance in which reforms to improve preparation practices can be developed; and (4) direct evidence of the extent to which teacher candidates demonstrate specific knowledge and skills expected by TPP faculty and staff (Peck, Singer-Gabella, Sloan, & Lin, 2014). Essentially, locally-scored performance assessments represent a promising measure for

evidence-based program improvement.

Despite this promise, TPPs can best rely on evidence from locally-scored performance assessments when the scores: (1) measure the constructs that they were designed to measure (construct validity); (2) are reliably scored by different individuals (reliability); and (3) predict teacher candidates' performance as classroom teachers (predictive validity) (Admiraal, Hoeksma, van de Kamp, & van Duin, 2011). Extant research suggests that teacher candidate performance assessments, like TPA, can be the fulcrum that leverages an evidence-based culture; however, without data that are valid, reliable, and predict outcomes of interest, the evidence provided by locally-scored performance assessments may not guide TPPs to adopt more effective preparation practices (Peck & McDonald, 2014; Peck Gallucci, Sloan, & Lippincott, 2009).

Therefore, for this study, we partnered with the College of Education at a large public university in North Carolina (hereon referred to as Collaborating University) to evaluate the construct validity, reliability, and predictive validity of their locally-scored performance assessment portfolios. Collaborating University (CU) used the widely-adopted TPA that was developed by Stanford University and is aligned with standards for TPPs (e.g. CAEP standards) and practicing teachers (e.g. the Interstate Teacher Assessment and Support Consortium, InTASC, standards). While the edTPA has recently replaced the TPA (SCALE, 2013), this study makes three contributions to the teacher candidate performance assessment research literature. First, this study focuses on the relationship between performance assessment scores and outcomes for program graduates—entry into and exit from the profession, teacher evaluation ratings, and teacher value-added scores. Second, this study compares local TPA portfolio scores to those from the official scorer, Pearson, to assess the utility of locally-scored measures as a guide for program improvements.² This is especially important given the centrality of local scoring in the current research on TPP reform and establishing a culture of evidence within TPPs (Miller, Carroll, Jancic, & Markworth, 2015; Peck, Gallucci, Sloan, & Lippincott, 2009, 2014; Peck & McDonald, 2014). Finally, this study serves as a proof of concept for the type of study that individual TPPs or collections of programs can undertake to establish the utility of local scoring of teacher candidate performance assessments to guide their own program improvement efforts. With 11 states requiring teacher candidate performance assessments for program completion and/or licensure decisions and over 600 universities using teacher candidate performance assessments, it is important to provide evidence on the validity and reliability of local scoring (edTPA, 2015).

In the sections that follow, we first provide further background on teacher candidate performance assessments. Specifically, we describe the origins of teacher candidate performance assessments and the organization of the TPA. Second, we detail CU's local scoring procedures, the TPA data and sample, and the outcome measures for the predictive validity analyses. Third, we present our analyses and findings. These analyses include more rigorous factor analysis models to assess construct validity, tests to assess the similarity of ratings from locally and officially-scored portfolios, and a range of regression models to determine whether local TPA scores predict teacher outcomes. Finally, we close with a discussion of the implications of our work for TPPs and their improvement efforts, policy action, and further research.

¹ In the recently released edTPA field test report, SCALE researchers reported two measures of inter-rater reliability: (1) the adjacent agreement rate and (2) the 'Kappa-N', which adjusts for inter-rater agreement by chance. Overall, these values were relatively high—0.917 and 0.829, respectively—and are comparable to reliability rates for other well-established performance assessments (SCALE, 2013).

² Pearson and its Evaluation Systems Group is a commercial education assessment organization that has partnered with SCALE to officially-score teaching candidates' edTPA portfolios.

Download English Version:

<https://daneshyari.com/en/article/6850438>

Download Persian Version:

<https://daneshyari.com/article/6850438>

[Daneshyari.com](https://daneshyari.com)