



## Research Memorandum

ETS RM-15-10

# Alignment Between the *Praxis*® Performance Assessment for Teachers (PPAT) and the Interstate Teacher Assessment and Support Consortium (InTASC) Model Core Teaching Standards

---

Clyde M. Reese

Richard J. Tannenbaum

Bamidele Kuku

October 2015

# ETS Research Memorandum Series

---

## EIGNOR EXECUTIVE EDITOR

James Carlson  
*Principal Psychometrician*

## ASSOCIATE EDITORS

Beata Beigman Klebanov  
*Senior Research Scientist – NLP*

Heather Buzick  
*Research Scientist*

Brent Bridgeman  
*Distinguished Presidential Appointee*

Keelan Evanini  
*Senior Research Scientist – NLP*

Marna Golub-Smith  
*Principal Psychometrician*

Shelby Haberman  
*Distinguished Presidential Appointee*

Donald Powers  
*Managing Principal Research Scientist*

Gautam Puhan  
*Principal Psychometrician*

John Sabatini  
*Managing Principal Research Scientist*

Matthias von Davier  
*Senior Research Director*

Rebecca Zwick  
*Distinguished Presidential Appointee*

## PRODUCTION EDITORS

Kim Fryer  
*Manager, Editing Services*

Ayleen Stellhorn  
*Editor*

---

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Memorandum series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Memorandum series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

**Alignment Between the *Praxis*<sup>®</sup> Performance Assessment for Teachers (PPAT)  
and the Interstate Teacher Assessment and Support Consortium (InTASC)  
Model Core Teaching Standards**

Clyde M. Reese, Richard J. Tannenbaum, and Bamidele Kuku  
Educational Testing Service, Princeton, New Jersey

October 2015

Corresponding author: C. Reese, E-mail: [CReese@ets.org](mailto:CReese@ets.org)

Suggested citation: Reese, C. M., Tannenbaum, R. J., & Kuku, B. (2015). *Alignment between the Praxis<sup>®</sup> Performance Assessment for Teachers (PPAT) and the Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards* (Research Memorandum No. RM-15-10). Princeton, NJ: Educational Testing Service.

Find other ETS-published reports by searching the ETS ReSEARCHER  
database at <http://search.ets.org/researcher/>

To obtain a copy of an ETS research report, please visit  
<http://www.ets.org/research/contact.html>

**Action Editor:** Heather Buzick

**Reviewers:** Joseph Ciofalo and Priya Kannan

Copyright © 2015 by Educational Testing Service. All rights reserved.

E-RATER, ETS, the ETS logo, and PRAXIS are registered trademarks of Educational Testing Service (ETS).

MEASURING THE POWER OF LEARNING is a trademark of ETS.

All other trademarks are the property of their respective owners.



**Abstract**

An alignment study was conducted with 13 educators who mentor or supervise preservice (or student teacher) candidates to explicitly document the connections between the Interstate Teacher Assessment and Support Consortium (InTASC) *Model Core Teaching Standards* and the *Praxis*<sup>®</sup> Performance Assessment for Teachers (PPAT). The multiple-task assessment requires candidates to submit written responses and supporting instructional materials and student work (i.e., artifacts). The PPAT was developed to assess a subset of the performance indicators delineated in the InTASC standards. In this study, we applied a multiple-round judgment process to identify which InTASC performance indicators are addressed by the tasks that compose the PPAT. The combined judgments of the experts determined the assignment of the InTASC performance indicators to the PPAT tasks. The panel identified 33 indicators measured by 1 or more PPAT tasks.

Key words: *Praxis*<sup>®</sup>, PPAT, InTASC, alignment

The interplay of subject-matter knowledge and pedagogical methods in the preparation and development of quality teachers has been a topic of discussion since the turn of the last century (Dewey, 1904/1964) and continues to drive the teacher quality discussion. Facilitated by the Council of Chief State School Officers (CCSSO), 17 state departments of education in the late 1980s began development of standards for new teachers that address both content knowledge and teaching practices (CCSSO, 1992). More recently, Deborah Ball and her colleagues have argued that “any examination of teacher quality must, necessarily, also grapple with issues of *teaching* quality” (Ball & Hill, 2008, p. 81). At the entry point into the profession—initial licensure of teachers—an added focus on the practice of teaching to augment subject-matter and pedagogical knowledge can provide a fuller picture of the profession of teaching.

The *Praxis*<sup>®</sup> Performance Assessment for Teachers (PPAT) is a multiple-task, authentic performance assessment completed during a candidate’s preservice, or student teaching, placement. The PPAT measures a candidate’s ability to gauge their students’ learning needs, interact effectively with students, design and implement lessons with well-articulated learning goals, and design and use assessments to make data-driven decisions to inform teaching and learning. The groundwork for the PPAT is the Interstate Teacher Assessment and Support Consortium (InTASC) *Model Core Teaching Standards and Learning Progressions for Teachers 1.0* (CCSSO, 2013). The multiple tasks within the PPAT address both (a) the separate components of effective practice and (b) the interconnectedness of these components. A multiple-round alignment study was conducted in February 2015 to explicitly document the connections between the InTASC standards and the PPAT. This report documents the alignment procedures and results of the study.

### **InTASC Standards and the PPAT**

The InTASC standards include 10 standards, and each standard includes performances, essential knowledge, and critical dispositions. For example, the first standard, Standard #1: Learner Development, includes three performances, four essential knowledge areas, and four critical dispositions (CCSSO, 2013). The PPAT focuses on a subset of the performances (referred to as *performance indicators*) as identified by a committee of subject-matter experts working with Educational Testing Service (ETS) performance assessment experts. The development of the PPAT began with defining a subset of the InTASC performance indicators (under the first nine standards<sup>1</sup>) that

- most readily applied to teacher candidates prior to the completion of their teacher preparation program (i.e., during preservice teaching),
- could be demonstrated during a candidate's preservice teaching assignment, and
- could be effectively assessed with a structured performance assessment.

The PPAT includes four tasks. Task 1 is a formative exercise and is locally scored; Task 1 does not contribute to a candidate's PPAT score. Tasks 2–4 are centrally scored and contribute to a candidate's score. Each task is composed of steps, and each step is scored using a unique, four-point scoring rubric. The step scores are summed to produce a task score—Task 2 includes three steps and the task-level score ranges from 3 to 12; Tasks 3 and 4 include four steps each and task-level scores range from 4 to 16. The task scores are weighted—the Task 4 score is doubled—and summed to produce the PPAT score. The current research addresses Tasks 2, 3, and 4, the three tasks that contribute to the summative, consequential PPAT score.

### **Alignment**

Alignment is typically considered as a component of content validity evidence that supports the intended use of the assessment results (Kane, 2006). Alignment evidence can include the connections between (a) content standards and instruction, (b) content standards and the assessment, and (c) instruction and the assessment (Davis-Becker & Buckendahl, 2013). While the content standards being examined are national in scope and the assessment was developed for national administration, the instruction provided at educator preparation programs (EPPs) across the country cannot be considered common. Therefore, connections with instruction are outside the scope of this research and attention was focused on the connection between the content standards—the InTASC standards—and the assessment—the PPAT.

Typically for licensure or certification testing, the content domain is defined by a systematic job or practice analysis (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The current InTASC standards were first published in 2011 (CCSSO, 2011) and were later augmented to include learning progressions for teachers (CCSSO, 2013). The InTASC standards have been widely accepted and were thus considered a suitable starting point for the development of the PPAT. The relevance and importance of the knowledge and skills contained

in the standards is supported by the literature on teaching (see the literature review commissioned by CCSSO at [www.ccsso.org/intasc](http://www.ccsso.org/intasc)).

To evaluate the content validity of the PPAT for the purpose of informing initial licensure decisions, evidence should be collected regarding relevance of the domain and alignment of the assessment to the defined domain (Sireci, 1998). As stated previously, the content domain for the PPAT is a subset of the performance indicators included in the InTASC standards. The initial development process, the recent steps to update the standards, and the research literature supporting the standards provides evidence of the strength of these standards as an accepted definition of relevant knowledge and skills needed for safe and effective teaching (CCSSO, 2013). Therefore, evidence exists to address the relevance and importance of the domain.

The purpose of this study is to explicitly evaluate the alignment of the PPAT to the InTASC standards to determine which of the InTASC standards and performance indicators are being measured by the three summative tasks that compose the PPAT. A panel of teacher preparation experts were charged with identifying any and all InTASC performance indicators that were addressed by the tasks. The combined judgments of the experts determined the assignment of the InTASC performance indicators to the PPAT tasks. Establishing the alignment of the tasks and rubrics to the intended InTASC performance indicators provides evidence to support the content validity of the PPAT. Content validity is critical to the proper use and interpretation of the assessment (Bhola, Impara, & Buckendahl, 2003; Davis-Becker & Buckendahl, 2013; Martone & Sireci, 2009).

### **Procedures**

A judgment-based process was used to examine the domain representation of the PPAT. The study took 2 days to complete. The major steps for the study are described in the following sections.

#### **Reviewing the PPAT**

Approximately 2 weeks prior to the study, panelists were provided with available PPAT materials, including the tasks, scoring rubrics, and guidelines for preparing and submitting supporting artifacts. The materials panelists reviewed were the same materials provided to candidates. Panelists were asked to take notes on tasks or steps within tasks, focusing on what



was being measured and the challenge the task poses for preservice teachers. Panelists also were sent the link to the InTASC standards and asked to review them.

At the beginning of the study, ETS performance assessment specialists described the development of the tasks and the administration of the assessment. Then, the structure of each task—prompts, candidate’s written response, artifacts, and scoring rubrics—were described for the panel. The whole-group discussion focused on what knowledge/skills were being measured, how candidates responded to the tasks and what supporting artifacts were expected, and what evidence was being valued during scoring.

### **Panelists’ Judgments**

The following steps were followed for each task. The panel completed all judgments for a task before moving to the next task. The panel received training on each type of judgment, the associated rating scale, and the data collection process. The judgment process started with Task 2 and was repeated for Tasks 3 and 4. The committee did not consider Task 1.

**Round 1 judgments.** The panelists reviewed the task and judged, for each step within the task, what InTASC standards were being measured by the step. The panelists made their judgments using a five-point scale ranging from 1 (*not measured*) to 5 (*directly measured*). InTASC standards that received a 4 or 5 by at least seven of the 13 panelists were considered measured by the task and thus considered in Round 2.

**Round 2 judgments.** For the InTASC standards identified in Round 1, the panelists judged how relevant each performance indicator under that standard was to successfully completing the step. For example, InTASC Standard #1: *Learner Development* has three performance indicators. The panelists made their judgments using a five-point scale ranging from 1 (*not at all relevant*) to 5 (*highly relevant*). Judgments were collected and summarized. InTASC performance indicators with an average judgment at or above 4.0 were considered aligned to the step.

**Round 3 judgments.** Next, the panel reviewed the rubric for each step and judged if the scoring rubric associated with the step addressed the performance indicators identified in Round 2. Based on the description of a candidate’s performance that would warrant the highest score of 4, the panel judged (“yes” or “no”) if the scoring rubric addressed the skills described in the performance indicator.

**Relevance, importance, and authenticity judgments.** Finally, the panelists indicated their level of agreement with the following statements:

- The skills being measured are relevant for a beginning teacher.
- The skills being measured are important for a beginning teacher.
- The task/step is authentic (e.g., represents tasks a beginning teacher can expect to encounter).

### **Final Evaluations**

The panelists completed an evaluation form at the conclusion of the study addressing the quality of the implementation and their certainty with their individual alignment judgments.

### **Results**

Alignment judgments, as well as relevance, importance and authenticity judgments, are summarized in the following sections.

#### **Round 1 Judgments**

The results from Round 1 (standards-level judgments) are a preliminary step to inform Rounds 2 and 3. To assure that all InTASC standards that may have some connection to a step were considered in Round 2, panelists' judgments were discussed and panelists could revisit their Round 1 judgments. Table 1 summarizes the Round 1 results.

**Table 1. Round 1 Alignment (Standard Level) Results**

PPAT task & step	Number of standards	Standards
Task 2/Step 1	5	1, 2, 6, 7, 8
Task 2/Step 2	5	1, 2, 6, 8, 9
Task 2/Step 3	5	1, 2, 6, 7, 9
Task 3/Step 1	7	1, 2, 3, 4, 5, 7, 8
Task 3/Step 2	6	1, 2, 4, 6, 7, 8
Task 3/Step 3	9	1, 2, 3, 4, 5, 6, 7, 8, 9
Task 3/Step 4	9	1, 2, 3, 4, 5, 6, 7, 8, 9
Task 4/Step 1	9	1, 2, 3, 4, 5, 6, 7, 8, 9
Task 4/Step 2	8	1, 2, 3, 4, 5, 6, 7, 8
Task 4/Step 3	6	1, 2, 4, 6, 7, 8
Task 4/Step 4	8	1, 2, 3, 4, 6, 7, 8, 9

*Note.* Task 1 was not judged by the reviewers and is not included in this table.

## **Round 2 Judgments**

Based on the results from Round 1, the panelists made alignment judgments for each performance indicator under the identified InTASC standards. Judgments were made using a five-point scale. Tables 2–4 summarize the Round 2 judgments for Tasks 2, 3, and 4, respectively. The shaded values indicate the performance indicators that met the criteria for alignment: mean judgment at or above 4.0 on the five-point scale. Only performance indicators meeting the criteria for alignment for one or more steps are included in the tables.

Given the strong interconnections among steps within a task and the reporting of candidate scores at the task level, the alignment of the PPAT to the InTASC standards is most appropriate at the task level. If a performance indicator is determined to be aligned to one or more steps, then it is aligned to the task. Table 5 summarizes the task-level alignment results from Round 2. The panel identified 33 performance indicators as being measured by one or more PPAT tasks.

## **Round 3 Judgments**

Based on the results from Round 2, the panelists made yes/no judgments regarding if the step-level rubric addressed each identified performance indicator. In all cases, a majority of the panelists indicated that the identified performance indicator was addressed by the step-specific rubric.<sup>2</sup> For all but eight of the 127 Round 3 judgments collected, more than 75% of panelists indicated the performance indicator was addressed; the judgment was unanimous for 56 of the step-indicators pairings.

## **Relevance, Importance and Authenticity of Tasks**

For each of the 11 steps that compose Tasks 2–4, the panelists<sup>3</sup> indicated their level of agreement with the following three statements:

- The skills being measured are relevant for a beginning teacher.
- The skills being measured are important for a beginning teacher.
- The task/step is authentic (e.g., represents tasks a beginning teacher can expect to encounter).

Tables 6–8 summarize the relevance, importance, and authenticity judgments.

**Table 2. Round 2 Alignment (Indicator Level) Results: Task 2**

Performance indicator <sup>a</sup>	Step 1	Step 2	Step 3
	Mean (SD)	Mean (SD)	Mean (SD)
1(a)	3.62 (1.04)	3.38 (1.19)	4.00 (1.22)
2(b)	4.15 (0.90)	3.46 (1.39)	3.62 (1.33)
2(f)	4.15 (0.90)	1.77 (0.83)	2.15 (0.99)
6(b)	4.54 (0.78)	2.23 (1.30)	2.92 (1.04)
6(c)	2.62 (1.45)	4.54 (0.97)	4.23 (1.09)
6(d)	2.00 (1.22)	4.08 (1.19)	2.15 (1.21)
6(g)	3.23 (1.36)	4.00 (1.08)	3.92 (1.12)
6(h)	4.31 (1.11)	3.69 (1.25)	3.08 (1.32)
7(d)	3.54 (1.13)	—	4.15 (1.28)
8(b)	3.23 (1.30)	4.15 (0.99)	—
9(c)	—	3.85 (1.28)	4.08 (1.26)

*Note.* Shaded values indicate performance indicators that met the criteria for alignment: mean judgment at or above 4.0 on the 5-point scale. As indicated by a dash, not all standards were identified in Round 1 judgments; therefore, Round 2 judgments were not collected for some performance indicators.

<sup>a</sup> Only performance indicators meeting the criteria for alignment for one or more steps are included.

**Table 3. Round 2 Alignment (Indicator Level) Results: Task 3**

Performance indicator <sup>a</sup>	Step 1	Step 2	Step 3	Step 4
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
1(a)	2.85 (1.46)	4.23 (1.17)	4.54 (0.52)	4.77 (0.44)
1(b)	4.85 (0.38)	4.85 (0.38)	4.15 (1.14)	4.08 (0.95)
2(a)	4.46 (0.66)	4.85 (0.38)	4.31 (0.95)	4.69 (0.63)
2(b)	4.23 (1.17)	4.77 (0.60)	4.46 (0.66)	4.54 (0.66)
2(c)	4.15 (1.34)	3.85 (1.14)	3.38 (1.33)	4.23 (1.01)
2(f)	4.08 (0.64)	3.69 (1.18)	3.54 (1.45)	4.31 (0.75)
3(e)	3.08 (1.61)	—	4.00 (1.29)	3.23 (1.48)
4(e)	4.08 (0.64)	3.31 (1.32)	3.46 (1.45)	3.92 (1.19)
4(f)	4.00 (1.08)	4.38 (0.51)	4.31 (0.63)	4.15 (0.55)
4(g)	4.31 (0.75)	3.54 (1.20)	3.77 (1.36)	3.85 (1.14)
6(a)	—	3.69 (1.03)	4.31 (0.85)	4.15 (1.21)
6(c)	—	3.77 (1.54)	4.31 (0.63)	4.54 (0.52)
6(d)	—	2.46 (1.33)	4.00 (1.15)	3.08 (1.38)
6(g)	—	4.31 (1.11)	4.00 (0.91)	4.00 (1.29)
7(a)	4.85 (0.38)	4.77 (0.44)	3.62 (1.71)	4.31 (0.85)
7(b)	5.00 (0.00)	4.77 (0.44)	3.92 (1.26)	4.46 (0.97)
7(c)	4.23 (1.01)	4.38 (0.65)	4.15 (1.34)	3.92 (1.19)
7(d)	4.38 (0.87)	4.31 (1.18)	3.77 (1.30)	4.54 (0.66)
7(f)	3.54 (1.45)	4.08 (1.12)	4.31 (0.95)	4.46 (0.52)
8(a)	4.92 (0.28)	4.92 (0.28)	4.69 (0.48)	4.69 (0.63)
8(b)	3.15 (1.52)	4.31 (1.11)	4.62 (0.51)	4.85 (0.38)
9(c)	—	—	4.15 (1.21)	(0.85)

*Note.* Shaded values indicate performance indicators that met the criteria for alignment: mean judgment at or above 4.0 on the five-point scale. As indicated by a dash, not all standards were identified in Round 1 judgments; therefore, Round 2 judgments were not collected for some performance indicators.

<sup>a</sup> Only performance indicators meeting the criteria for alignment for one or more steps are included.

**Table 4. Round 2 Alignment (Indicator-Level) Results: Task 4**

Performance indicator <sup>a</sup>	Step 1 Mean (SD)	Step 2 Mean (SD)	Step 3 Mean (SD)	Step 4 Mean (SD)
1(a)	4.62 (0.51)	4.62 (0.65)	4.77 (0.44)	4.69 (0.63)
1(b)	4.69 (0.63)	3.77 (1.48)	3.85 (1.41)	3.69 (1.32)
2(a)	4.62 (0.65)	4.23 (1.30)	4.23 (1.17)	4.15 (0.90)
2(b)	4.23 (1.17)	3.85 (1.34)	4.00 (1.29)	3.85 (1.14)
2(c)	4.54 (0.78)	3.23 (1.36)	3.46 (1.39)	3.54 (1.45)
3(d)	3.77 (1.01)	4.46 (0.66)	—	3.38 (1.26)
3(f)	3.08 (1.38)	4.69 (0.48)	—	2.92 (1.38)
4(c)	3.62 (1.04)	4.00 (1.22)	2.54 (1.51)	2.92 (1.38)
4(d)	4.00 (1.08)	3.92 (1.12)	2.69 (1.49)	2.92 (1.26)
4(f)	4.00 (1.08)	3.92 (1.26)	3.69 (1.55)	4.08 (1.12)
4(h)	4.15 (0.99)	3.69 (1.25)	2.15 (1.21)	2.54 (1.05)
5(h)	4.62 (0.51)	4.62 (0.51)	—	—
6(a)	4.69 (0.48)	4.23 (1.09)	4.23 (1.09)	4.15 (1.21)
6(b)	4.46 (0.97)	3.62 (1.45)	4.23 (1.17)	3.62 (1.39)
6(c)	3.92 (1.26)	3.85 (1.21)	4.69 (0.48)	4.15 (1.28)
6(g)	4.15 (1.07)	4.00 (1.08)	4.23 (1.09)	4.23 (0.73)
7(a)	4.85 (0.38)	4.08 (1.32)	3.85 (1.34)	4.00 (1.22)
7(b)	4.77 (0.44)	4.15 (1.34)	3.92 (1.12)	4.38 (0.77)
7(c)	4.31 (1.11)	3.85 (1.41)	3.46 (1.33)	3.54 (1.33)
7(d)	4.69 (0.48)	3.77 (1.24)	3.92 (1.26)	4.38 (0.87)
7(f)	3.92 (1.12)	3.54 (1.33)	3.23 (1.30)	4.54 (0.66)
8(a)	4.38 (1.12)	4.15 (1.46)	3.46 (1.13)	4.38 (0.77)
8(b)	4.69 (0.48)	4.85 (0.38)	4.23 (1.17)	4.69 (0.63)
8(f)	4.38 (0.65)	4.46 (0.52)	2.31 (1.44)	2.92 (1.19)
8(h)	4.46 (0.78)	4.54 (0.52)	3.00 (1.53)	3.31 (1.25)
8(i)	4.62 (0.51)	4.54 (0.52)	2.46 (1.56)	2.92 (1.26)
9(c)	3.92 (1.19)	—	—	—

*Note.* Shaded values indicate performance indicators that met the criteria for alignment: mean judgment at or above 4.0 on the five-point scale. As indicated by a dash, not all standards were identified in Round 1 judgments; therefore, Round 2 judgments were not collected for some performance indicators.

<sup>a</sup> Only performance indicators meeting the criteria for alignment for one or more steps are included.

**Table 5. Round 2 Task-Level Alignment (Indicator Level) Results**

PPAT task	Number of indicators	Indicators
Task 2	11	1(a), 2(b), 2(f), 6(b), 6(c), 6(d), 6(g), 6(h), 7(d), 8(b), 9(c)
Task 3	22	1(a), 1(b), 2(a), 2(b), 2(c), 2(f), 3(e), 4(e), 4(f), 4(g), 6(a), 6(c), 6(d), 6(g), 7(a), 7(b), 7(c), 7(d), 7(f), 8(a), 8(b), 9(c)
Task 4	27	1(a), 1(b), 2(a), 2(b), 2(c), 3(d), 3(f), 4(c), 4(d), 4(f), 4(h), 5(h), 6(a), 6(b), 6(c), 6(g), 7(a), 7(b), 7(c), 7(d), 7(f), 8(a), 8(b), 8(f), 8(h), 8(i), 9(c)
Overall	33	1(a), 1(b), 2(a), 2(b), 2(c), 2(f), 3(d), 3(e), 3(f), 4(c), 4(d), 4(e), 4(f), 4(g), 4(h), 5(h), 6(a), 6(b), 6(c), 6(d), 6(g), 6(h), 7(a), 7(b), 7(c), 7(d), 7(f), 8(a), 8(b), 8(f), 8(h), 8(i), 9(c)

*Note.* Task 1 was not judged by the reviewers and is not included in this table.

**Table 6. Relevance, Importance, and Authenticity Judgments: Task 2**

Indicator	Step	Strongly agree		Agree		Disagree		Strongly disagree	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Relevance	1	9	69	4	31	0	0	0	0
Importance	1	7	54	6	46	0	0	0	0
Authenticity	1	5	38	7	54	1	8	0	0
Relevance	2	7	54	6	46	0	0	0	0
Importance	2	7	54	6	46	0	0	0	0
Authenticity	2	6	46	6	46	1	8	0	0
Relevance	3	7	54	6	46	0	0	0	0
Importance	3	8	62	5	38	0	0	0	0
Authenticity	3	5	38	7	54	1	8	0	0

**Table 7. Relevance, Importance and Authenticity Judgments: Task 3**

Indicator	Step	Strongly agree		Agree		Disagree		Strongly disagree	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Relevance	1	10	77	3	23	0	0	0	0
Importance	1	11	85	2	15	0	0	0	0
Authenticity	1	10	77	3	23	0	0	0	0
Relevance	2	10	77	3	23	0	0	0	0
Importance	2	11	85	2	15	0	0	0	0
Authenticity	2	9	69	4	31	0	0	0	0
Relevance	3	8	62	4	31	1	8	0	0
Importance	3	10	77	2	15	1	8	0	0
Authenticity	3	7	54	4	31	2	15	0	0
Relevance	4	7	54	6	46	0	0	0	0
Importance	4	10	77	3	23	0	0	0	0
Authenticity	4	7	54	4	31	2	15	0	0

**Table 8. Relevance, Importance and Authenticity Judgments: Task 4**

Indicator	Step	Strongly agree		Agree		Disagree		Strongly disagree	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Relevance	1	9	69	4	31	0	0	0	0
Importance	1	9	69	4	31	0	0	0	0
Authenticity	1	8	62	4	31	1	8	0	0
Relevance	2	9	69	4	31	0	0	0	0
Importance	2	8	62	5	38	0	0	0	0
Authenticity	2	8	62	4	31	1	8	0	0
Relevance	3	8	62	5	38	0	0	0	0
Importance	3	8	62	5	38	0	0	0	0
Authenticity	3	8	62	5	38	0	0	0	0
Relevance	4	5	50	5	50	0	0	0	0
Importance	4	6	60	4	40	0	0	0	0
Authenticity	4	4	40	6	60	0	0	0	0

*Note.* Ten of the 13 panelists completed Round 3 judgments for Task 4/Step 4.

For each of the steps across Tasks 2, 3 and 4, all or all but one of the panelists agreed or strongly agreed that the skills being measured are relevant and important for beginning teachers. Except for two steps, all or all but one of the panelists agreed or strongly agreed the activities were authentic; 11 of the 13 panelists agreed or strongly agreed for Steps 3 and 4 of Task 3.

### **Sources of Evidence Supporting the Alignment**

In discussing the evidence supporting the results of the PPAT-InTASC alignment study, material will be organized based on the framework presented by Davis-Becker and Buckendahl (2013) for evaluating alignment studies. Based on a similar framework presented by Kane (2001) for evaluating standard-setting studies, the framework includes

- procedural evidence (description of panel and panelists' evaluations),
- internal evidence (consistency of judgments),
- external evidence (consistency with developers' judgments, InTASC progressions),  
and
- utility evidence (input to ongoing development).

The following discussion focuses on procedural, internal, and external evidence; all results from the study and feedback from panelists were shared with the assessment development team to inform ongoing development of the PPAT and similar performance assessments (utility evidence).

Given that validity is an accumulation of evidence rather than a yes/no determination, structuring the discussion by these components will allow test users, as well as the test provider, to evaluate and interpret the study's results in light of the intended uses of the PPAT scores.

### **Procedural Evidence**

The literature agrees that the panelists must be familiar with the content standards (i.e., InTASC standards) and the target population for the test (Davis-Becker & Buckendahl, 2013). The panelists should also be independent of the development process so as not to have a conflict of interest (Webb, 1999; Bhola et al., 2003). However, the literature is less consistent regarding the size of an alignment study panel, with panel sizes as small as two reported for some methodologies (Porter, 2002). Webb (2007) recommended panels of between five and eight

panelists, but the upper limit is actually set by the need for diversity among panelists and the capacity of the facilitator to manage effective training and meaningful discussion.

The multistate alignment panel was composed of 13 educators from eight states (Arkansas, Maryland, Mississippi, Nebraska, New Jersey, North Carolina, Pennsylvania, and West Virginia) and Washington, DC. All the educators were involved with the preparation and supervision of prospective teachers. The majority of panelists (11 of the 13 panelists) were college faculty or associated with a teacher preparation program; the remaining two panelists worked in K–12 school settings. All the panelists reported mentoring or supervising preservice, or student, teachers in the past 3 years. Finally, all 13 panelists indicated they were at least somewhat familiar with the InTASC standards; approximately half (seven of the 13 panelists) indicated they were very familiar (see Table 9).

**Table 9. Panelists Background**

Characteristic	<i>N</i>	%
Current position		
K–12 teacher	2	15
Administrator	1	8
College faculty	10	77
Gender		
Female	10	77
Male	3	23
Race		
White	6	46
Black or African American	3	23
Hispanic or Latino	1	8
Asian or Asian American	2	15
Other	1	8
Mentored or supervised preservice teachers in the past 3 years		
Yes	13	100
No	0	0
Experience mentoring or supervising preservice teachers		
3 years or less	2	15
4–9 years	3	23
10–14 years	2	15
15 years or more	6	46
No experience	0	0
Familiarity with InTASC <i>Model Core Teaching Standards</i>		
Not familiar	0	0
Somewhat familiar	6	46
Very familiar	7	54



Selection of appropriate methodology and assembling a panel of subject-matter experts are critical first steps in planning and conducting a sound alignment study. However, it is critical that the panelists are well trained in the methodology and are prepared to make informed judgments. At the conclusion of the 2-day study, panelists indicated their level of agreement to three statements regarding the training:

- I understood the purpose of this study.
- The facilitator's instructions and explanations were clear.
- The facilitator's instructions and explanations were easy to follow.

Panelists also answered three statements regarding their familiarity with the PPAT and the InTASC standards:

- I understood the InTASC standards well enough to make my judgments.
- I understood the PPAT tasks/steps well enough to make my judgments.
- I understood the PPAT rubrics well enough to make my judgments.

Finally, the panelists were asked how certain they were with their alignment judgments.

Overall, panelists felt well trained for the judgment exercises; all panelists agreed or strongly agreed that they understood the purpose of the study and that instructions/explanations were clear and easy to follow. All the panelists also agreed or strongly agreed that they understood the InTASC standards, the PPAT tasks/steps, and the step-specific rubrics well enough to complete their judgments. Finally, all the panelists reported they were certain or very certain of the judgments they made during the study.

### **Internal Evidence**

In Round 2, panelists made 534 step-indicator judgments using a 5-point rating scale. One approach to examining the consistency of the panel's judgments is to examine the standard error of judgment (SEJ) for each step-indicator pairing. The SEJ is the standard deviation of the panelists' judgments divided by the square root on the number of panelists (Cizek & Bunch, 2007).

Across tasks, 85% of the 534 step-indicator pairings had an SEJ less than or equal to 0.40 (or 10% of the range of a five-point rating scale). Only one of the SEJs for the 127 aligned step-indicator pairings was greater than 0.40.

**External Evidence**

The alignment methodology employed in this study relied on the informed judgments of subject-matter experts (panelists) who reviewed both the InTASC standards and the PPAT tasks and rubric. The panelists were not involved in the development of the PPAT. Two additional points of reference for evaluating the results of the alignment study are (a) classifications of tasks by the assessment specialists during the development of the PPAT and (b) the learning progressions developed by the consortium (CCSSO, 2013).

**Consistency with developers' classifications.** During the development process, ETS assessment specialists, working with a committee of subject-matter experts with qualifications similar to the study's panelists, identified the performance indicators measured by each PPAT task. The criteria for "measured" was intentionally permissive to cast a wide net. Performance indicators that were only tangentially measured by the task were identified.

The panel of subject-matter experts identified 11 performance indicators for Task 2, 22 for Task 3, and 27 for Task 4. As described previously, the classification criteria applied during the development of the PPAT set a lower bar for attaching a performance indicator to a task; therefore, slightly more indicators were identified during development. Comparing the panel's results with the developers' classifications, 82% (9 of 11) of the identified indicators matched for Task 2, 90% (18 of 21) matched for Task 3, and 85% (23 of 27) matched for Task 4.

**InTASC progressions.** As part of the revisions to the InTASC standards in 2013, the consortium included learning progressions for teachers throughout their professional lifespan. The progressions "articulate a continuum of growth and higher levels of performance" (CCSSO, 2013, p.10) for teachers throughout their career trajectory. The standards are cross-walked to the descriptive text of the each progression (three levels are described). Performance indicators (as well as essential knowledge and critical dispositions) can appear in more than one of the three progression levels. The application of a performance indicator would increase in complexity and sophistication as a teacher progresses through the levels.

The three progression levels accompanying the InTASC standards were intentionally not named to avoid the label assigned to restricting teaching performance. However, it can be assumed that Level 1, the lowest level, would include preservice teachers and teachers just entering the profession. Of the 64 performance indicators under Standards 1–9, nearly three-quarters, or 49 indicators, initially appeared under the first progression level, though these

indicators may have appeared in higher levels also. The remaining 15 indicators first appeared in a later level.

Given the test-taking audience for the PPAT—preservice teachers—the tasks would be most appropriate if measuring those indicators that would most likely fall in the first learning progression level. As shown in Table 5, 33 performance indicators were identified as aligning to PPAT Tasks 2–4. Thirty of the 33 aligned indicators initially appeared under the first learning progressions level. The remaining indicators—Indicators 3(e), 6(h), and 7(b)—initially appeared in the second level.

### **Conclusions**

The PPAT was designed to be aligned to the InTASC standards and to serve as a measure of teaching quality. The PPAT would be a component of a state’s initial licensure system and would be administered during a candidate’s preservice (or student teaching) placement. Candidates’ submit written responses and supporting instructional materials and student work (i.e., artifacts) to demonstrate their ability to gauge their students' learning needs, interact effectively with students, design and implement lessons with well-articulated learning goals, and design and use assessments to make data-driven decisions to inform teaching and learning.

The InTASC standards include 10 standards and each standard includes performances, essential knowledge, and critical dispositions. The PPAT focuses on a subset of the performances (referred to as “performance indicators”) as identified by a committee of subject-matter experts working with ETS assessment experts. The current study identified the InTASC performance indicator measured by the three PPAT tasks that contribute to the overall, consequential score. Overall, 33 performance indicators were identified as being measured by one or more of the tasks (see Table 5).

In addition to the alignment of the PPAT tasks to the InTASC standards, panelists also judged the relevance and importance of the skills being measured for beginning teachers and the authenticity of the tasks. For each step within the tasks, the skill being measured were judged to be relevant and important for beginning teachers. The steps/tasks also were judged to be authentic (e.g., represent tasks a beginning teacher can expect to encounter).

### References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Ball, D. L., & Hill, H. C. (2008). Measuring teacher quality in practice. In D. H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (pp. 80–98). Thousand Oaks, CA: Sage.
- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational measurement: Issues & practices*, 22, 21–29.
- CCSSO. (1992). *Model standards for beginning teacher licensing, assessment and development: A resource for state dialogue*. Retrieved from <http://programs.ccsso.org/content/pdfs/corestrd.pdf>
- CCSSO. (2011). *InTASC model core teaching standards: A resource for state dialogue*. Retrieved from [http://www.ccsso.org/Documents/2011/InTASC\\_Model\\_Core\\_Teaching\\_Standards\\_2011.pdf](http://www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf)
- CCSSO. (2013). *InTASC model core teaching standards and learning progressions for teachers 1.0*. Retrieved from <http://programs.ccsso.org/content/pdfs/corestrd.pdf>
- Cizek, G. J., & Bunch, M. (2007). *Standard setting: A practitioner's guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Davis-Becker, S. L., & Buckendahl, C. W. (2013). A proposed framework for evaluating alignment studies. *Educational measurement: Issues & practice*, 32(1), 23–33.
- Dewey, J. (1964). The relation of theory to practice in education. In R. Archambault (Ed.), *John Dewey on education* (pp. 313–338). Chicago, IL: University of Chicago Press. (Original work published in 1904.)
- Kane, M. T. (2001). So much remains the same: Conceptions and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53–88). Mahwah, NJ: Erlbaum.
- Kane, M. T. (2006). Validation. *Educational measurement*, 4, 17–64. Westport, CT: Praeger.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of educational research*, 79(4), 1332–1361.

- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14.
- Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, 5, 299–321.
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7–25.

### **Notes**

<sup>1</sup> Given the intended audience for the assessment, Standard 10: Leadership and Collaboration, was not included in the PPAT.

<sup>2</sup> For several indicators aligned to Task 2 and one indicator aligned to Task 4, a data collection error resulted in not collecting Round 3 judgments. Also, for Task 4/Step 4, Round 3 judgments were collected for 10 of the 13 panelists.

<sup>3</sup> For Task 4, Step 4, 10 of the 13 panelists completed this portion of their judgments.