



Development of a Measure of Classroom Assessment Practices for Certification of Accomplished Teachers

Carol Ezzelle*

National Board for Professional Teaching Standards, Arlington, VA, United States

OPEN ACCESS

Edited by:

Sarah M. Bonner,
Hunter College (CUNY), United States

Reviewed by:

Giray Berberoglu,
Başkent University, Turkey
Shenghai Dai,
Washington State University,
United States

*Correspondence:

Carol Ezzelle
cezzelle@nbpts.org

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 26 April 2019

Accepted: 06 August 2019

Published: 10 September 2019

Citation:

Ezzelle C (2019) Development of a
Measure of Classroom Assessment
Practices for Certification of
Accomplished Teachers.
Front. Educ. 4:90.
doi: 10.3389/feduc.2019.00090

Teachers pursuing National Board Certification complete a rigorous assessment comprised of four components measuring standards of accomplished pedagogical knowledge and practice. One of those components of the assessment, *Component 4: Effective and Reflective Practitioner*, is a new performance-based portfolio component that centers on the use of assessment data and other sources of information to advance student and teacher growth. Specifically, the component requires teachers to:

- gather and utilize data from different sources to build their knowledge and understanding of students;
- base instructional decision-making and assessment practices on their knowledge of students, collaboration with stakeholders, and adherence to assessment principles; and
- use their knowledge of students to reflect on their practice, advance their professional learning, and promote student growth throughout learning communities.

The development and outcomes of the pilot test of *Component 4: Effective and Reflective Practitioner* related to establishing validity evidence for the component's scores as a measure of accomplished teaching are described. The methods reported cover the use of experts in the development of instructions, a pilot study, and formative scoring to build evidence of the validity based on test content and ensure the successful launch of the new component.

Keywords: certification, classroom assessment, test development, validity, assessment literacy, accomplished teaching, teaching standards

INTRODUCTION

At the root of the National Board for Professional Teaching Standards' mission is the establishment of the profession's standards for accomplished teaching and providing a national system for certifying those who meet those standards as National Board Certified Teachers (NBCTs). The basis upon which teachers are assessed, National Board Standards require teaching practices that ensure students meet high, college- and career-ready standards of learning. National Board Certification is the profession's mark of accomplished teaching, an endorsement from practicing peers who developed the Standards and assessment and, as such, the National Board program is "by teachers, for teachers." The assessment along with the standards they measure are the cornerstone of the National Board's strategy for building a coherent career continuum that ensures all teachers have an early vision of accomplished teaching, deliberately build their practice to an accomplished level, and continue to grow and lead as accomplished practitioners.

Since first offering certification in 1994, the National Board has established that its assessment reliably certifies teachers who meet accomplished standards. More than 122,000 teachers in 25 certificate areas have achieved the title of National Board Certified Teacher, and a growing body of evidence continues to show that Board-certified teachers have a positive impact on achievement based on increases in students' test scores (National Research Council, 2008; Chingos and Peterson, 2011; Strategic Data Project, 2012a,b; Cavalluzzo et al., 2014; Cowan and Goldhaber, 2015; National Board for Professional Teaching Standards, 2016a). Fulfilling its responsibility as a certifying body, the National Board first reconfigured its assessment in 2000 by streamlining six portfolio components down to four. In 2013 the National Board began a redesign of its entire assessment with the intention to incorporate advances in research, practice, and measurement while understanding that the choices made have implications for the validity of score interpretation and use. Ongoing involvement of Board-certified teachers in the redesign guaranteed that the assessment remained relevant to teachers, capturing characteristics of accomplished practice as defined by National Board Standards.

Anchored within the Five Core Propositions (National Board for Professional Teaching Standards, 2016b) and informed by National Board Standards in 25 certificate areas, the peer-reviewed assessment asks teachers to demonstrate their practice by showing what they know and do. For each certificate area, the assessment comprises four components: one computer-based component and three portfolio components. Each component is briefly described below and details can be found at <http://www.nbpts.org/national-board-certification>.

Component 1: Content Knowledge

This computer-based assessment measures a teacher's understanding of content knowledge and pedagogical practices for teaching the certificate-specific content area. These domains are assessed through the completion of constructed response exercises and selected response items.

Component 2: Differentiation in Instruction

This classroom-based portfolio entry requires that the teacher gather and analyze information about his/her individual students' strengths and needs and use that information to design and implement instruction to advance student learning and achievement. A teacher submits work samples that demonstrate the students' growth over time and a written commentary that analyzes his/her instructional choices.

Component 3: Teaching Practice and Learning Environment

This is a classroom-based portfolio entry that requires video recordings of interactions between the teacher and his/her students. The teacher also submits a written commentary in which he/she describes, analyzes and reflects on his/her teaching and interactions with students and the impact on their learning.

Component 4: Effective and Reflective Practitioner

This portfolio entry requires the teacher to demonstrate evidence of his/her abilities as an effective and reflective practitioner in developing and applying knowledge of his/her students; his/her use of assessments to effectively plan for and positively impact his/her students' learning; and his/her collaboration to advance students' and teachers' learning and growth.

In order to earn National Board Certification, a teacher must earn scores at or above (1) a minimum score on the section that measures content and pedagogical practice (Component 1), (2) a minimum average score across portfolio-based components (Components 1, 2, and 3), and (3) a required total score. More details on scoring can be found in the *Scoring Guide: Understanding Your Scores* (National Board for Professional Teaching Standards, 2019).

The majority of the assessment redesign resulted in changes to the way evidence of accomplished teaching was collected, e.g., by adding selected response items to the computer-based component, or by combining two video-based components into one component. The redesign also included the replacement of a component with a newly developed component, *Component 4: Effective and Reflective Practitioner*. This paper presents the development of that component and how its development, including iterative reviews of instructions and scoring rubrics by subject matter experts and the pilot testing of the component by Board-eligible teachers, contributed to evidence of the new component's content validity.

MATERIALS AND METHODS

Component Design

The design of the new component involved deliberations and decisions of multiple groups of subject matter experts at different stages. While this paper focuses on the development of one new component of the assessment, the National Board initially gathered three committees of specialists from across the educational field to begin the redesign of the entire certification assessment (National Board for Professional Teaching Standards, 2016c). These individual expert committees possessed the technical and pedagogical expertise to make substantive recommendations on the measurement of current preK-12 teaching practices and learning outcomes. The committees deliberated at length regarding research trends and their potential application to the assessment of accomplished teaching that is defined by the National Board Standards. They considered issues related to the collection and analysis of meaningful evidence, and they made design suggestions for all components of the certification assessment.

Second, a separate group of subject matter experts, the assessment review panel, studied these suggestions based on operational requirements and real-world demands. Both the committees and the assessment review panel concluded that a new portfolio component evaluating the assessment literacy of teachers would strengthen the certification process while advancing the field's understanding of how to measure

teacher impact on student learning across all grade levels and subject areas.

Third, to initiate development of the new component, an additional committee representing a wide range of expertise defined assessment and data literacy within the scope of National Board Certification by drafting a test construct stating that *Component 4: Effective and Reflective Practitioner* measures teachers' understanding and management of data and assessments to effectively support their students' growth. The component's content domain included learning goals and the purposes of assessment; selection, implementation, and analysis of assessments; collaboration and communication about assessment and student learning; student involvement in assessment; and reflective use of assessment. As such, teacher evidence of demonstrating the multi-faceted Component 4 construct would include development of a student group profile based on various data they collected; demonstration of knowledge of sound assessment principles and how they used the information gained from assessments and other data sources to positively impact these students' learning and become better teachers.

Fourth, NBCTs then mapped the relationship among all components of the assessment and the foundational Five Core Propositions and the National Board Standards for each of the 25 certificate areas. This crosswalk guided subsequent work by ensuring that Component 4 was well-aligned with the comprehensive tenets of accomplished teaching and complemented the structure and content of other components. This crosswalk revealed the need and opportunity to capture evidence of the ways in which teachers use professional interactions with colleagues, families, and local communities to compliment data and inform their judgements about student learning.

Last, the procedures to arrive at the Component 4 design and the design itself were supported by the Technical Advisory Group in adherence with criteria established by the Standards for Educational and Psychological Testing (American Educational Research Association (AERA), 2014).

Component Instructions, Assessment Tasks, and Scoring Rubric

For the first step of developing the Component 4 instructions, assessment tasks, and scoring rubric, a model document was created based on the construct and content domain. Once the model document was created, subject matter experts were recruited to judge the appropriateness and content representation of the materials. A content advisory committee whose members included 57 NBCTs from 16 certificate areas within 11 disciplines was culled from a pool of applicants who responded to a survey that self-assessed their assessment and data literacy. To be eligible for the content advisory committee, an applicant must have reported the following: self-reported ratings of *advanced* or *expert* as his/her current depth of knowledge/understanding of assessment and data literacy, recent professional development in the area of assessment, experience developing and using holistic and analytic scoring rubrics, experience using data to inform instruction, and classroom practice providing their students with assessment feedback.

The specific purpose of the content advisory committee was to review the model set of materials which maintained the same structure/approach for all certificate areas and provide feedback on its suitability for their specific certificate area. Guiding review questions included: is content aligned with the certificate area-specific standards? Is content and scope of entry sufficient to distinguish accomplished from not-yet-accomplished practice? Does the scope of requirements seem reasonable? The committee's input allowed for adjustments to the instructions, assessment tasks and scoring rubric to be made both globally across certificate areas and in specific certificate areas as needed before pilot testing the component with Board-eligible teachers. Input from measurement experts was also utilized as the National Board met with its Technical Advisory Group periodically to monitor development and discuss the psychometric implications of methods to bolster the evidence of validity and reliability of the assessment's scores.

The bulleted information below briefly summarizes the assessment tasks required for Component 4. For full documents for each certificate area, visit <https://www.nbpts.org/national-board-certification/candidate-center/first-time-and-returning-candidate-resources/>.

- **Contextual Information.** Submit a form that describes the broader context in which you teach, e.g., the type of school/program in which you teach, the grade/subject configuration, and the number of students and courses you teach.
- **Knowledge of Students.** Select one class or group of students as the focus for both the Knowledge of Students and the Generation and Use of Assessment Data sections of this portfolio entry. Submit a completed Group Information and Profile Form and associated evidence.
- **Generation and Use of Assessment Data.** Select two assessments—one formative and one summative—to use in this portfolio entry. Submit the following forms that describe these assessment materials:
 - **Instructional Context Form**
 - **Formative Assessment Materials Form** and associated evidence, including the assessment or a description of it, assessment results, and student self-assessments
 - **Summative Assessment Materials Form** and associated evidence.
- **Participation in Learning Communities.** Describe a professional learning need and a student need that you have met by working collaboratively with colleagues or about which you have shared your expertise in a leadership role with the larger learning community. Submit the following forms that describe these needs:
 - **Description of Professional Learning Need Form** and associated evidence
 - **Description of a Student Need Form** and associated evidence
- **Written Commentary.** Write a commentary on your practice of gathering and using information about students and how you contribute to positive changes for students.

Pilot Test

The pilot test approach encouraged widespread participation among Board-eligible teachers while furthering final improvements to component instructions, assessment tasks, and scoring rubrics. The pilot test did not purport to examine the psychometric properties of the new component but instead intended to demonstrate the feasibility of teachers' completion and the scoring of the new component. A subset of NBCTs from the content advisory committee served as advisors to one to six pilot test participants in their certificate area. The advisor's role was only to provide general guidance, e.g., directing participants to pilot test resources. Support of the pilot test approach was rooted in the engagement of NBCTs in the actual development of the component's instructions, tasks scoring rubric and its scoring, thus strengthening evidence of content validity.

Participants

Practicing teachers who were pursuing National Board Certification were recruited to participate in the pilot test. While many pilot participants were recruited through a web site application, some of the teachers were recruited by content advisory committee members. Written and informed consent was received from all participants via an agreement that specified services provided by participants, their compensation, and assignment of ownership of all work to National Board. An ethics approval was not required as per applicable institutional and national guidelines and regulations.

As an incentive, teachers completing the pilot test were granted the right to resubmit portions of their Component 4 portfolios for certification purposes. Other incentives to participate included a certificate of participation, a letter of commendation, an honorarium of \$500, and an entry into a drawing for \$1,000. Pilot test participants were emailed approximately every 2 weeks with content encouraging them to progress through the sections of Component 4.

The pilot test for Component 4 was conducted in 11 of the 17 disciplines available. These certificate areas were selected to represent core subject areas, a range of grade levels, non-classroom teacher positions, and varied instructional configurations (e.g., English as a new language, exceptional needs specialist), disciplines that had unique considerations (e.g., career and technical education, school counseling, library media specialist), and/or disciplines having sizeable literature on issues specific to assessing teacher practice, e.g., the performing arts.

Sample size for Component 4 pilot test was targeted at 176 (16 participants in each of the 11 disciplines). At the beginning of the pilot test window, there were 212 pilot test participants. Ultimately, 139 (66%) participants submitted their Component 4 portfolio. **Table 1** presents the number of participants who submitted a Component 4 portfolio by certificate area. Nine of 11 disciplines had a sample size <16, which may indicate the challenges that participants might have had in completing this performance-based component in a timeframe of just over 3 months.

As for demographics, 91% of pilot test participants were female; 76% were white, 13% African American, 3% Asian, 3% Hispanic, and 4% unknown ethnicity. Forty-five percent of the participants reported 3–10 years teaching experience, 34%

TABLE 1 | Number of component 4 pilot participants by certificate area.

Certificate area	N
Art	11
Early Adolescence through Young Adulthood	7
Early and Middle Childhood	4
School Counseling/Early Childhood through Young Adulthood	15
Career and Technical Education/Early through Young Adulthood Adolescence	7
English Language Arts	16
Adolescence Young Adulthood	8
Early Adolescence	8
English as a New Language	12
Early and Middle Childhood	10
Early Adolescence through Young Adulthood	2
Exceptional Needs Specialist/Early Childhood through Young Adulthood	11
Generalist/Early Childhood	19
Library Media/Early Childhood through Young Adulthood	15
Math	10
Adolescence Young Adulthood	6
Early Adolescence	4
Music	11
Early and Middle Childhood	9
Early Adolescence through Young Adulthood	2
Physical Education	12
Early and Middle Childhood	8
Early Adolescence through Young Adulthood	4
Total	139

reported more than 10 but <20 years' experience, and 16% reported 20 years or more experience. The participants were located across rural (19%), suburban (32%), and urban (27%) district locations (unknown for 21%). These demographics of the pilot test sample closely aligned with those of the population of teachers pursuing National Board Certification; however the pilot study purpose was not generalization but rather to demonstrate that the completion and scoring of the new component was feasible and provided evidence of content validity.

Pilot Test Participant Survey

A survey that asked for perceptions and feedback about the component instructions, assessment tasks, and rubrics was emailed to the 139 participants who submitted portfolios. The survey included nine questions with Yes/No response options (seven of those prompted further information if the response was No) and one question that asked for an estimate of how many hours they spent on developing their portfolio.

Formative and Pilot Scoring

For scoring the pilot test submissions, 60 assessors were recruited spanning the 11 disciplines in the pilot test. Each discipline had three to eight assessors. Some assessors were NBCTs with experience in scoring National Board's assessment, and others were NBCTs from the content advisory committee. Assessors used the National Board's 12-point rubric score scale which is based on four primary levels of performance (Levels 4, 3, 2, and 1), with plus (+) and minus (-) variations at each level. The highest assigned score for a constructed response item or a portfolio component is 4.25 (4+); the lowest is

0.75 (1–) (National Board for Professional Teaching Standards, 2019). Level 4 and Level 3 performances, or score of 2.75 or higher, represent accomplished teaching practice. Level 2 and Level 1 performances represent less-than-accomplished teaching practice. Zero scores were earned if a critical piece of evidence (e.g., written commentary) was missing from the submitted component portfolios.

The task of formative scoring focused on four pre-selected Component 4 pilot submissions that represented a range of quality and approaches. Grouped by discipline, assessors reviewed each submission independently, one at a time, and discussion followed each submission. Next, pilot scoring included independent scoring, discussion, and assignment of a consensus score for the remainder of the pilot test submissions. Simultaneously, the committees re-evaluated the Component 4 instructions, tasks and rubrics in light of the pilot test submissions.

At the most general level, the purpose of the pilot scoring activities was to confirm and/or improve the quality of the assessment instructions, assessment tasks, and scoring rubrics. Several procedures were employed to accomplish this goal with the intent to provide evidence to support claims that National Board wished to make about the component. These goals, claims, and activities were an important component of the validity argument process.

The goal of confirming and/or improving the quality of the component materials can be stated as the specific goals of the pilot scoring process:

- determine the scorability of the responses provided to the assessment tasks, and
- review and refine the Component 4 instructions, assessment tasks and rubrics.

The pilot scoring activities were designed to support several content validity claims associated with these goals. These claims stated that the component materials:

1. are aligned with the Five Core Propositions associated with the National Board Standards;
2. are stated in clear, understandable, and unambiguous language;
3. are equally accessible by teachers across certificate areas; and
4. elicit responses that contain relevant and scorable evidence regarding whether a teacher exhibits accomplished levels of content knowledge and teaching practices.

The specific activities that were undertaken during the pilot scoring process were designed to support these claims and goals, and they included:

1. assessors pre-scoring followed by group discussion resulting in holistic consensus scores to pilot submissions;
2. assessors completing a survey that contained items related to scorability; and,
3. refinement of the Component 4 instructions, assessment tasks, and rubrics through expert review.

RESULTS

Scorability

Evidence of scorability of the Component 4 materials was reflected in the assignment of scores to all submissions and in survey results from the assessors. The procedure of consensus scoring resulted, by its nature, in agreement among assessors about the scores to be assigned. Consensus scoring did not allow for agreement statistics or reliability estimates. Statistics

TABLE 2 | Component 4 score statistics and distributions across certificate areas.

Certificate area	N	Mean	Min	Max	Percentage in score range*							
					(0, 0.75)	[0.75, 1.25]	(1.25, 1.75)	[1.75, 2.25]	(2.25, 2.75)	[2.75, 3.25]	(3.25, 3.75)	[3.75, 4.25]
Generalist/Early Childhood	19	2.342	1.00	4.00	0.0	21.1	0.0	31.6	0.0	36.8	0.0	10.5
School Counseling	15	2.450	1.00	4.00	0.0	13.3	0.0	40.0	0.0	33.3	0.0	13.3
Library Media	15	2.383	1.00	4.00	0.0	6.7	0.0	60.0	0.0	26.7	0.0	6.7
Exceptional Needs Specialist	11	2.364	1.00	4.00	0.0	9.1	0.0	54.5	0.0	27.3	0.0	9.1
English as a New Language/EMC	10	2.300	1.00	4.00	0.0	30.0	0.0	20.0	0.0	40.0	0.0	10.0
Music/EMC	9	2.333	1.00	4.00	0.0	44.4	0.0	11.1	0.0	11.1	0.0	33.3
English Language Arts/AYA	8	2.125	1.00	4.00	0.0	25.0	0.0	50.0	0.0	12.5	0.0	12.5
English Language Arts/EA	8	1.625	0.00	3.00	12.5	25.0	0.0	50.0	0.0	12.5	0.0	0
Physical Education/EMC	8	2.688	2.00	4.00	0.0	0.0	0.0	37.5	12.5	37.5	0.0	12.5
Art/EAYA	7	2.250	1.25	4.00	0.0	28.6	0.0	42.9	0.0	14.3	0.0	14.3
Career and Technical Education	7	1.286	0.00	3.00	28.6	28.6	0.0	28.6	0.0	14.3	0.0	0
Mathematics/AYA	6	2.167	0.00	3.00	16.7	0.0	0.0	33.3	0.0	50.0	0.0	0.0
Mathematics/EA	4	1.313	0.00	3.25	50.0	0.0	0.0	25.0	0.0	25.0	0.0	0.0
Physical Education/EAYA	4	1.750	0.00	3.00	25.0	0.0	0.0	50.0	0.0	25.0	0.0	0.0
Art/EMC	4	2.000	1.00	3.00	0.0	25.0	0.0	50.0	0.0	25.0	0.0	0.0
English as a New Language/EAYA	2	1.875	1.75	2.00	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Music/EAYA	2	1.875	1.75	2.00	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0

*Brackets indicate inclusion of a score; parentheses indicate values up to a particular score.

AYA, Adolescence and Young Adulthood; EA, Early Adolescence; EAYA, Early Adolescence through Young Adulthood; EMC, Early and Middle Childhood.

for Component 4 scores were computed for each certificate area and are reported in **Table 2**. Due to the small sample sizes in all certificate areas, caution should be used when interpreting statistics. In addition, the likelihood of lower levels of motivation and preparedness among pilot test participants compared with operational candidates should be considered.

The range of scores declined as the number of submissions scored within a certificate area decreased from a maximum of 19 (Generalist/Early Childhood) to a minimum of two (English as a New Language/ and Music/ Early Adolescence through Young Adulthood). The percentage of scores at 2.75 or higher that indicate clear evidence of accomplished teaching in a certificate area ranged from 0% (for the two certificate areas with only two teachers) to 50%. Nine of 17 (53%) certificate areas had seven to 33% of scores at Level 4 (3.75 or higher) indicating the submissions demonstrated clear, consistent, and convincing evidence of accomplished teaching. Despite small samples sizes, the observed range of scores reinforced the expectation that the scoring differentiated across levels of performance.

Scorability was also assessed by multiple questions in a survey completed by the assessors upon completion of scoring (**Table 3**). Eighty-five percent of assessors said the rubric clearly described expected performance at each score level, and 88% said the various parts of the component worked in concert so that

a single holistic score could be assigned. Seventy-five percent did not find any issues with the wording of the score-level descriptions in the rubric that might negatively affect scorers' understanding of the score levels. Only 37% saw submissions with evidence types that are not easily evaluated with the rubric. The assessors provided suggestions to improve the scoring rubric and instructions in an effort to improve the scorability of the new component.

Refinement of Instructions, Assessment Tasks, and Scoring Rubrics

The refinement of the *Component 4: Effective and Reflective Practitioner* instructions, assessment tasks and scoring rubric was informed by feedback from both the pilot test participants who developed and submitted the portfolios and the assessors who scored the materials. Just as during development of the materials, information was again collected on the appropriateness and content representation of the component instructions and tasks to establish evidence of content validity. The feedback from the assessors and participants ensured that materials were revised based on very specific feedback.

Participant Survey Results

A total of 110 of the 139 (79%) participants responded to the survey on perceptions and feedback about the component

TABLE 3 | Component 4 pilot test assessors' (N = 60) survey results.

Questions	Response	N	%
Are the tasks aligned with the Five Core Propositions of the National Board?	No	0	0
	Yes	60	100
Do the tasks seem appropriate for the certificate and developmental level?	No	0	0
	Yes	60	100
Do the activities associated with the portfolio entry, as a whole, represent authentic practices in the field of focus?	No	5	8
	Yes	55	92
Does the required evidence provide a clear platform for the demonstration of accomplished teaching?	Missing	1	–
	No	8	14
	Yes	51	86
Are the tasks described in a manner that is clear and easy to understand?	Missing	1	–
	No	36	61
	Yes	23	39
Are there any issues with the wording of the directions that might negatively affect candidates' understanding of the tasks or the way they organize their cases?	Missing	1	–
	No	10	17
	Yes	49	83
Does the rubric clearly describe expected performance at each score level?	No	9	15
	Yes	51	85
Are there any issues with the wording of the score-level descriptions in the rubric that might negatively affect scorers' interpretation or understanding of the score levels?	No	45	75
	Yes	15	25
In light of the cases reviewed, do the tasks work in concert so that relative case strengths and weaknesses can be weighed (via the rubric) and a single holistic score can be assigned?	No	7	12
	Yes	53	88
While reviewing the pilot cases, did you see any on-task cases with approaches or evidence types that are not easily evaluated with the rubric?	No	38	63
	Yes	22	37
While reviewing the pilot cases, did you see any cases that relied on evidence that seemed more manufactured for the assigned tasks than authentic to the volunteer's teaching assignment?	No	35	58
	Yes	25	42
Do the page limits allow sufficient room for candidates to address the tasks?	No	3	5
	Yes	57	95

TABLE 4 | Component 4 pilot test participants' ($N = 110$) survey results.

Questions	Response	%
Were the instructions and expectations for this entry clear and easy to understand?*	No	40
	Yes	60
Were the forms associated with this portfolio entry clear and easy to understand?*	No	26
	Yes	74
Were the questions to guide the Written Commentary clear and comprehensive?*	No	13
	Yes	87
Did the rubric clearly describe expected performance?	No	9
	Yes	91
Was the overall focus of the portfolio entry clear?*	No	20
	Yes	80
Did you find that the allotted space and page counts were sufficient?*	No	28
	Yes	72
Do you feel that the information and evidence required for this entry gave you adequate opportunity to demonstrate your professional skills and abilities in the areas of focus of this entry (ex: knowledge of students, generation and use of data, participation in learning communities)?*	No	20
	Yes	80
Did the activities and tasks associated with the portfolio entry, as a whole, represent authentic practices in your field?*	No	18
	Yes	83
Did the time and effort to plan, collect evidence, and complete the entire portfolio entry seem reasonable?	No	26
	Yes	74

*If the participant's response was No, the survey prompted for elaboration and suggestions for improvement.

instructions, tasks and rubrics. Results for nine of the 10 questions are summarized in **Table 4**. The tenth question asked for an estimate of time spent developing their portfolio materials; 65% of pilot test participants reported they spent 10–60 h and 35% reported 61 h or more. Twenty-six percent stated the time and effort to complete the portfolio did not seem reasonable; these results could reflect that the pilot test window (February–May) was shorter than an operational window where teachers can begin as early as April the year before.

Despite the majority of pilot test participants responding affirmatively to the clarity of the component materials and opportunity to demonstrate their skills in their area, there was still a significant percentage who responded otherwise which indicated a need for improvement of the materials. For the seven survey questions on clarity and adequacy of materials, and the question of task authenticity, a follow-up question asked for suggestions for improvement of the materials. Pilot test participants' suggestions were used to refine component instructions, tasks and rubrics across all certificate areas.

Assessor Survey Results

At the conclusion of scoring all submissions, the 60 assessors provided survey feedback on the component regarding whether the instructions elicited the kind of responses anticipated and recommended additional changes to the instructions to make them ready for operational use. Responses are summarized in **Table 4**. Compared to pilot participants, a greater percentage (61%) of assessors indicated that the instructions were not easy to understand; revisions based on assessors' suggestions were made accordingly. In support of validity based on test content,

all assessors agreed the tasks were aligned with the Five Core Propositions and were appropriate for the certificate area.

Final Review

After implementing feedback from pilot test participants and assessors, the final certificate-specific instructions, tasks, and scoring rubric for piloted and non-piloted areas were reviewed by 22 NBCTs representing all 11 pilot tested disciplines and who were involved in prior reviews and/or scoring. Overall, the feedback from the NBCTs was positive and indicated that the revisions made following the pilot test improved the clarity of the instructional materials. No additional substantive revisions were suggested by these NBCT reviewers, and the materials were finalized for operational use.

DISCUSSION

The education profession has long known that changes to the classroom assessment environment would require examination of the changing roles and responsibilities of the teacher along with the impact of understanding and skills required to develop and implement assessment effectively in the classroom (Stiggins, 1991). National Board was proactively attentive to current research, and engaged educational specialists, technical advisors, and NBCTs as the agents of innovation and operationalization through the development of instructions, assessment tasks, and scoring rubric; administration of a pilot study; and formative and pilot scoring of a new component of its assessment of accomplished teaching used for National Board Certification. The methods used in the development and piloting of *Component 4: Effective and Reflective Practitioner* met the

goals of utilizing various subject matter input to refine the component instructions, tasks and scoring rubric, improve its scorability and assess the component's fidelity with the Five Core Propositions and Standards thereby establishing evidence of validity for its use as a measure of classroom assessment practices. By doing so, National Board was able to readily and with ease operationalize the new component that evaluates classroom practice, professional learning, and collaboration, examining the impact that a teacher's assessment and data literacy has in student learning outcomes.

National Board Certification remains "by teachers, for teachers" as the profession's mark of accomplished teaching, and its goals and objectives were strengthened by updates in structure and methodology using best practices in test development. As *Component 4: Effective and Reflective Practitioner* moves into its third year of operational administration, nearly 20,000 teachers have submitted a Component 4 portfolio on their journey to National Board Certification. The 2018 interrater agreement rates, defined as the percentage of two assessors' scores that are ≤ 1.25 points apart on the rubric scale, range from 94.2 to 100%. The future direction of research is aimed at accumulating more validity evidence for this component as a measure of a teacher's carefully coordinated instructional and assessment practices supporting the growth and development of their students, their colleagues, and themselves.

REFERENCES

- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Cavalluzzo, L., Barrow, L., Henderson, S., Mokher, C., Geraghty, T., and Sartain, L. (2014). *From Large Urban to Small Rural Schools: An Empirical Study of National Board Certification and Teaching Effectiveness*. Arlington, VA: CNA Education. Retrieved from: http://www.nbpts.org/sites/default/files/Policy/cna_from_large_urban_to_small_rural_schools.pdf
- Chingos, M. M., and Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: familiar and new results on the correlates of teacher effectiveness. *Econ. Educ. Rev.* 30, 449–465. doi: 10.1016/j.econedurev.2010.12.010
- Cowan, J., and Goldhaber, D. (2015). *National Board Certification and Teacher Effectiveness: Evidence from Washington*. CEDR Working Paper 2015-3. Seattle, WA: University of Washington.
- National Board for Professional Teaching Standards (2016a). *Research*. Retrieved from: <http://www.nbpts.org/advancing-education-research>
- National Board for Professional Teaching Standards (2016b). *What Teachers Should Know and Be Able to Do, 2nd Edn*. Arlington, VA: National Board for Professional Teaching Standards. Retrieved from: http://www.nbpts.org/sites/default/files/what_teachers_should_know.pdf
- National Board for Professional Teaching Standards (2016c). *National Board Certification: A Redesign Framework for Performance-Based Assessments of Accomplished Teaching*. Unpublished manuscript, Arlington, VA: National Board for Professional Teaching Standards.

DATA AVAILABILITY

The datasets for this study will not be made publicly available because they are property of National Board for Professional Teaching Standards.

AUTHOR CONTRIBUTIONS

The author confirms being the main contributor of this brief research report and approved it for publication.

FUNDING

This work in this paper was supported in part by the Bill and Melinda Gates Foundation (Grant Number OPP1088311).

ACKNOWLEDGMENTS

I thank my colleagues, many from National Board and Pearson, who provided their expertise and efforts to the work cited herein, including but not limited to Trey Clifton, Andrea Hajek, Kristin Hamilton, Cathy Hawks, Ryan Lankford, Susan Lopez Bailey, Lisa Stooksberry, and Melissa Villareal. National Board Certified Teachers nationwide receive my gratitude.

- National Board for Professional Teaching Standards (2019). *Scoring Guide: Understanding and Interpreting Your Scores*. Retrieved from: https://www.nbpts.org/wp-content/uploads/NBPTS_Scoring_Guide.pdf
- National Research Council (2008). *Assessing Accomplished Teaching: Advanced-Level Certification Programs*. Washington, DC: The National Academies Press.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan* 72, 534–539.
- Strategic Data Project (2012a). *Learning About Teacher Effectiveness: SDP Human Capital Diagnostic, Gwinnett County Public Schools, GA*. Cambridge, MA: Center for Education Policy Research; Harvard University. Retrieved from: [http://www.gse.harvard.edu/\\$\sim\\$pfpie/pdf/gwinnett_human-capital_may2012.pdf](http://www.gse.harvard.edu/\simpfpie/pdf/gwinnett_human-capital_may2012.pdf)
- Strategic Data Project (2012b). *SDP Human Capital Diagnostic: Los Angeles Unified School District*. Cambridge, MA: Center for Education Policy Research; Harvard University. Retrieved from: <http://www.gse.harvard.edu/~pfpie/pdf/sdp-laUSD-hk-brief.pdf>

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ezzelle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.