

Re-examining teacher candidate performance on world language edTPA as a consequential assessment

Susan A. Hildebrandt, *Illinois State University*

Pete Swanson, *Georgia State University / United States Air Force Academy*

Abstract

Used by over 800 educator preparation programs in more than 40 states and the District of Columbia, edTPA seeks to measure content-specific outcomes of beginning teachers as a high-stakes assessment. Many states also use edTPA to inform teacher licensure or certification decisions, making the assessment consequential for individual teacher candidates. The present study focuses on teacher candidates' performance on the World Language edTPA, targeting two world language teacher education programs since the assessment became consequential for licensure in their respective states. It then compares findings to those of a previous study of the same programs' World Language edTPA performance (Hildebrandt & Swanson, 2014) from before the assessment became consequential for teacher licensure decisions in 2015 in Illinois and Georgia. Programmatic changes were implemented after initial inspection of the data, and results from such curricular changes are discussed in terms of local impact. Additionally, national trends are identified and discussed. Findings demonstrate

Susan A. Hildebrandt (Ph. D., The University of Iowa) is Professor of Applied Linguistics and Spanish at Illinois State University, where she coordinates a world language teacher education program and regularly teaches world language pedagogy classes. Her research concerns world language teacher knowledge, development, evaluation, and professionalization and has been published in *Hispania*, *Foreign Language Annals*, and *Teaching and Teacher Education*.

Pete Swanson (Ph.D., University of Wyoming) is Professor of Foreign Language Education at Georgia State University where he serves as Coordinator of the world language teacher education program. Currently, Dr. Swanson is serving as Distinguished Visiting Professor of Spanish at the United States Air Force Academy in Colorado. His research focuses on teacher effectiveness and the identification, recruitment, and retention of world language teachers.

NECTFL Review, Number 84, September 2019, pp. 15–33. © 2019 by Northeast Conference on the Teaching of Foreign Languages.

serious issues regarding this assessment of beginning teacher effectiveness as related to World Language teaching and learning. This research has implications for teacher preparation programs and other educational stakeholders alike.

At the time of writing, 834 Educator Preparation Programs in 41 states and the District of Columbia (American Association for Colleges of Teacher Education, AACTE, 2018a) are participating in edTPA, a high-stakes performance assessment of new teacher readiness. Its widespread use for graduation and licensure decisions has prompted concerns among teacher educators and teacher candidates alike. In general, teacher educators worry that their programs will be evaluated by their teacher candidates' edTPA scores and that their teacher candidates may not be able to pass edTPA. Teacher candidates worry about the assessment's cost and that their score may be insufficient for graduation and licensure.

... edTPA, a high-stakes performance assessment of new teacher readiness. Its widespread use for graduation and licensure decisions has prompted concerns among teacher educators and teacher candidates alike.

Over the last several years, edTPA has become an integral part of the student teaching experience for tens of thousands of student teachers. Many states have set minimum cut scores that teacher candidates must meet to earn licensure or certification in the state, with necessary pass scores ranging from 29 (of 65 points possible) in Oregon to 37 in Illinois beginning September 2019 (AACTE, 2018b). Other states, like Minnesota, use edTPA scores to evaluate teacher preparation programs (AACTE, 2018b), while some universities require that teacher candidates pass edTPA for graduation from a teacher education program. Scores are frequently used to gather data that can promote curricular changes to teacher education programs and content modifications for methods courses that prepare teacher candidates. As opposed to other frequently used multiple-choice assessments of teacher knowledge (e.g., state content exams, Praxis II World Language Pedagogy test), edTPA requires teacher candidates to actively demonstrate their pedagogical knowledge by submitting artifacts from a series of lessons taught to actual students in the form of a lengthy portfolio.

The widespread use of edTPA across the country prompted the current exploration of edTPA scores in two teacher education programs in order to compare early results (Hildebrandt & Swanson, 2014) to those after September 1, 2015, when edTPA became required for licensure in Illinois and Georgia. The programs' scores will also be compared to national World Language (WL) edTPA averages.

Literature Review

In this literature review, the researchers explore previous studies of WL edTPA candidate performance and national trends, criticisms of edTPA, and the conceptual framework of interventions undertaken by the two target WL teacher education programs after an initial pilot study.

Re-examining teacher candidate performance on world language edTPA

WL edTPA and national trends. This study serves as a follow-up to the original exploratory investigation of edTPA prior to its consequential use in Illinois and Georgia (Hildebrandt & Swanson, 2014). That study compared teacher candidates' composite scores with known edTPA cut scores in the early-adopting states of Washington and New York, finding that all participants would have passed edTPA in Washington, with its cut score of 30, and 90% would have passed with New York's cut score of 35. In that study, the researchers also reported that the programs' teacher candidates performed best on the Planning task, second best on the Instruction task, and most poorly on the Assessment task. Behney (2018) confirmed the same pattern across six WL teacher preparation programs throughout the United States.

Other studies of WL edTPA have focused on the rhetorical demands of lengthy commentaries submitted by teacher candidates. As explained in Troyan and Kaplan's (2015) case study of one Spanish teacher candidate's evolution as a writer, the exigencies of the critical academic reflection genre required by edTPA commentaries necessitate candidates to "describe and analyze artifacts of practice" (p. 374). That description and analysis, provided by the commentaries for each task, explains the pedagogical decisions made by each candidate during the Planning, Instruction, and Assessment tasks. To alleviate the writing demands placed on candidates, Troyan and Kaplan (2015) called for explicit instruction during coursework to help move candidates' writing from personal private reflection to critical academic reflection.

These rhetorical challenges for native speakers of English may be exacerbated for non-native speakers of English. Russell and Davidson Devall (2016) found that, among their seven participants (of whom one was a native speaker and another was a heritage speaker of Spanish), a misalignment occurred between edTPA scores and other assessments, like evaluations from the university supervisor and cooperating teacher, content knowledge tests, and proficiency in the target language, as measured by the Oral Proficiency Interview. This misalignment was particularly profound for non-native English speakers, given the writing demands of the three edTPA commentaries. Supporting those findings, Jourdain (2018) reported that native and heritage speakers of Spanish submitted their portfolios for evaluation in smaller numbers than their native English-speaking classmates and that those non-native English speakers' WL edTPA scores were lower than those of their native English-speaking peers.

Mentor and university supervisors' familiarity, or lack thereof, with edTPA can also influence candidates' performance on the WL edTPA. Behney (2016) explored the critical role that cooperating teachers have on candidates' performance, finding that more experienced teachers play an integral role in teacher candidates' edTPA success. Russell and Davidson Devall (2016) also found that mentors' support levels may be hampered by a lack of knowledge of edTPA and how it fits in the larger teacher candidate evaluation landscape. In turn, that information deficit may influence teacher candidate edTPA performance.

Mentor and university supervisors' familiarity, or lack thereof, with edTPA can also influence candidates' performance on the WL edTPA.

National edTPA data trends show that from 2014 to 2016 (Stanford Center for Assessment, Learning, and Equity, SCALE, 2015, 2016a, 2017a), as the number of teacher candidates who took the WL edTPA increased from 416 to 655, the total mean scores declined a little over 10%, or 4.1 points. That is, total mean scores dropped from 40.0 in 2014 to 35.9 in 2016, with the means for all tasks decreasing. Of particular interest is the drop on Rubric 8 (Subject-Specific Pedagogy) in the Instruction task, which dropped from 2.4 to 1.9 over the course of those three years (SCALE, 2015, 2016a, 2017a). No rationale for lower scores over the course of the three administrations was given in the administrative reports.

Criticisms of edTPA. Various criticisms have been levied against edTPA in general and the WL edTPA in particular, although it is an improvement over many previously existing assessments of teacher readiness (Au, 2013). Some accuse edTPA of deprofessionalizing teaching and teacher education through the encroachment of corporations like Pearson into educational decision-making (Dover, Schultz, Smith, & Duggan, 2015; Madeloni & Gorlewski, 2013). Others argue that edTPA depersonalizes teaching and teacher education as it diminishes local control of teacher preparation (Cochran-Smith, Piazza, & Power, 2013) and teacher candidates’ attention to diversity as they prepare their edTPA lessons (Au, 2013). edTPA’s \$300 price tag is also a concern, as over 40,000 teacher candidates’ portfolios were expected for official scoring in 2017, creating a revenue stream of \$12 million (SCALE, 2016b). This cost contributes to the already high fees associated with graduating from teacher education programs and earning licensure in a state (Hildebrandt & Swanson, 2014, 2016).

WL edTPA has its own set of challenges living up to the professional expectations of the WL teaching profession (Hildebrandt & Swanson, 2016). Although it focuses on student second language proficiency development in a meaningful cultural context, stipulating a *meaningful* cultural context seems to ignore the fact that all human interactions occur within cultural contexts. What makes a cultural context meaningful? As for student proficiency development, none of the WL edTPA’s 13 rubrics focus on teacher use of the target language, which is an integral source of students’ comprehensible input. In the video recordings of the Instruction task, teacher candidate use of target language is not evaluated, despite the large amount of attention devoted to that skill in WL teacher education programs (American Council on the Teaching of Foreign Languages, ACTFL, 2013), ACTFL’s (2010) suggestion of 90% or more target language use, and the field’s general focus on students’ language proficiency development. Further, despite WL edTPA rubrics’ emphasis on function over form, Swanson and Hildebrandt (2017) reported that the portfolios of their teacher candidates with the highest Assessment task subscores contained a large number of objectives that focused on form instead of function.

Although it focuses on student second language proficiency development in a meaningful cultural context, stipulating a *meaningful* cultural context seems to ignore the fact that all human interactions occur within cultural contexts.

Re-examining teacher candidate performance on world language edTPA

Promoted as a content-specific assessment of teacher performance, edTPA portfolios are evaluated by scorers with “content-specific expertise” who are external to the teacher candidate’s home institution (Pearson Technology, 2018, para. 3). While the qualifications of those scorers focus on classroom teaching and supervisory experiences (Pearson Technology, 2018), without addressing teaching philosophies and whether scorers engage in up-to-date teaching practices those qualifications are vague (Swanson & Hildebrandt, 2017). That is, scorers who practice outdated pedagogy may be hired to evaluate edTPA portfolios based on their experiences teaching and supervising. The lack of transparency related to scorer qualifications challenge the WL teacher education community and teacher candidates who were taught in their methods classes to teach language in communicative ways that may not be rewarded in the evaluation of their portfolios. Further, the dearth of model WL edTPA portfolios for teacher educator and teacher candidate reference further disadvantage stakeholders by keeping secret the criteria that make an outstanding portfolio. Sample WL edTPA portfolios are scant in the field, so local training may be performed using portfolios from other content areas. Having WL faculty members and teacher candidates examine portfolios from other content areas is not as helpful as having them explore WL portfolios that can support an understanding of the real exigencies of the content-specific assessment.

Conceptual Framework of Interventions. Developing content-specific teacher knowledge extends across collegiate departmental lines (e.g., Department of Languages and Literatures, Department of Secondary Education). Coursework at the postsecondary level generally develops content knowledge in language classes housed in a language department, general pedagogical knowledge in a College of Education, and pedagogical content knowledge in WL pedagogy classes offered either in the College of Education or a language department, depending on the university context (Hildebrandt, Hlas, & Conroy, 2013). The WL pedagogy classes, often called methods, are critical to teacher candidate success on edTPA because of the content-specific nature of the assessment.

Developing content-specific teacher knowledge extends across collegiate departmental lines.

edTPA requires teacher candidates to enact teaching approaches that move beyond grammar-based methods. To succeed on the assessment, teacher candidates must apply communicative language teaching approaches to the creation of portfolios that demonstrate efforts to develop students’ communicative proficiency in meaningful cultural context(s) and to determine if students can, in fact, meet objectives set out in advance. They must also carry out that teaching while connecting the target language and culture(s) to students’ academic background knowledge and personal, cultural, and community assets (SCALE, 2017b).

ACTFL’s (2015) core practices were explored and implemented in the target programs’ methods classes, with the objective of improving scores in each of the three WL edTPA tasks. These core practices are critical to successful student language learning:

- Use target language for learning
- Design communicative activities

NECTFL Review 84

- Teach grammar as concept and use in context (IPA)
- Use authentic cultural resources
- Plan with backward design model
- Provide appropriate feedback

In particular, the use of Integrated Performance Assessments (IPA) was emphasized in order to boost Planning task subscores after it was found that previous teacher candidates who used an IPA in their edTPA learning segments tended to receive higher scores on their portfolios than those who did not (Swanson & Goulette, 2018). For the Instruction task, instructors highlighted the critical need to differentiate instruction for the diverse learners in the classroom and to remove barriers to learning as much as possible by using the principles of Universal Design (Scott, McGuire, & Foley, 2003).

Based on previous teacher candidates' performance (Hildebrandt & Swanson, 2014), the most attention was paid to improving subscores on the Assessment task. Backwards design (Wiggins & McTighe, 2005), in which teachers state learning outcomes and design assessments prior to planning instruction, played an integral role in WL methods classes at both universities. Determining objectives and student learning outcomes (Gronlund, 2004) before designing assessments and eventually lessons played a central role throughout the teacher preparation programs. These practices were explicitly modeled as part of the WL methods courses, with learning outcomes and their connection to the standards for teacher preparation (ACTFL, 2013) made for each in-class activity and all assignments.

These interventions, prompted by the 2014 study, informed the two research questions addressed in the present study:

1. What effect did interventions related to planning, instruction, and assessment in WL pedagogy classes have on teacher candidates' edTPA performance?
2. How do the data from these two programs compare to national WL edTPA averages?

Methods

This quantitative study investigates teacher candidate performance on the WL edTPA, as compared to a previous study of the same two WL teacher education programs' candidates' performance and national averages on the assessment.

Institutional Contexts. The authors are WL teacher education program coordinators at Illinois State University and Georgia State University, respectively. Although a Georgia State University faculty member, the second author is currently serving as a Distinguished Visiting Professor at the United States Air Force Academy. Illinois State University is located in Normal, Illinois and is a moderately large public institution with 20,635 students, with 72.6% White, 10.3% Hispanic, and 8.9% African American (Illinois State University, 2018). It is the oldest public university in the state and one of the largest producers of teachers in the nation (Illinois State University, 2019). Georgia State University is a public, urban research institution located in downtown Atlanta, Georgia with a student population of over 51,000 students (Georgia State University, 2019). It is one of the most diverse universities in the United States Georgia State University, 2019).

Re-examining teacher candidate performance on world language edTPA

While obvious differences between the two universities exist, multiple commonalities are also present. First, Illinois State University and Georgia State University have a combined enrollment of approximately 75 teacher candidates—a 65 percent drop in enrollment over the several years since the previous study (Hildebrandt & Swanson, 2014), which is consistent with national data on teacher education enrollment in general (Long, 2016). Teacher candidates at both institutions complete general pedagogical coursework in general foundations of education, technology integration, and working with diverse student populations, as well as six credits of coursework in pre-K to grade 12 WL pedagogy that centers on standards-based, proficiency-oriented approaches to instruction and assessment. At both universities, teacher candidates are placed for field experiences in a variety of diverse pre-K to grade 12 schools (e.g., rural, suburban, urban). Finally, both universities are regionally accredited and earned national recognition from the Council for the Accreditation of Educator Preparation. Due to such commonalities and given the relatively minor programmatic differences between the two, the two samples of teacher candidate scores on the WL edTPA were combined to form one data set.

Participants. Following Institutional Review Board approval, 36 teacher candidates in the two aforementioned WL teacher education programs agreed to participate in the research. The participants represent the total number of graduates during the five semesters from fall 2015 to fall 2017. All participants were seeking initial certification in Spanish or French education, and the majority of participants are female (70%). All participants created a WL edTPA portfolio during their final field placement (e.g., student teaching) and submitted it for official external review. All participants but two passed the WL edTPA with scores sufficient for licensure in their state on the first attempt; the two whose initial scores were insufficient passed on the second attempt. Per edTPA guidelines at both institutions regarding the filming of K-12 students, parents / guardians of the students in the participants' classrooms were asked to grant their permission to videotape their children and use the data for research purposes.

Instrument. Using the Common Core State Standards and the National Standards for Foreign Language Learning (National Standards for Foreign Language Education Project, 2006) as a framework, the WL edTPA was designed by the AACTE and Stanford University to “measure novice teachers’ readiness to teach world language” (SCALE, 2017b, p. 1) while grounded on a three-step teaching cycle of planning, instruction, and assessment. The developers of the WL edTPA state that the assessment is grounded in research and theory, as well as on findings that successful teachers

- develop knowledge of subject matter, content standards, and subject-specific pedagogy
- develop and apply knowledge of varied students’ needs
- consider research and theory about how students learn
- reflect on and analyze evidence of the effects of instruction on student learning. (p. 1)

Working independently of college faculty members, teacher candidates must plan a learning segment of three to five lessons, analyze their instruction via video clips that are submitted for review, and use student data to inform their practice.

NECTFL Review 84

Specifically, teacher candidates must submit via Pearson's online system lesson plans, copies of instructional and assessment materials, up to 15 minutes of video clips showing in-class instruction, and student work samples, justifying their pedagogical decisions via written commentaries. Materials such as the WL edTPA handbook, rubrics, and training/scoring materials are proprietary with exclusive authorship and copyrights held by SCALE; therefore, such specific information cannot be included here due to copyright restrictions. In states where edTPA is used for consequential purposes (e.g., licensure), the portfolios are evaluated externally by scorers trained by Pearson. However, during edTPA pilot testing, portfolios can be scored locally for formative purposes after having been trained by their local edTPA coordinator or by SCALE.

The WL edTPA, in particular, centers on students' communicative proficiency development in a meaningful cultural context. Portfolios are scored by current or retired K-12 faculty or administrators, teacher education faculty, or university supervisors (Pearson Technology, 2018). WL edTPA scores can range from zero to 65 total points, with 13 five-point Likert scale rubrics: Planning for Instruction and Assessment (Rubrics 1, 2, 3 and 4), Instructing and Engaging Students in Learning (Rubrics 5, 6, 7, 8 and 9), and Assessing Student Learning (Rubrics 10, 11, 12 and 13). Each rubric is scored from 1 to 5:

- Level 1 represents the low end of the scoring spectrum, representing the knowledge and skills of a struggling candidate who is not ready to teach;
- Level 2 represents the knowledge and skills of a candidate who is possibly ready to teach;
- Level 3 represents the knowledge and skills of a candidate who is ready to teach;
- Level 4 represents a candidate with a solid foundation of knowledge and skills for a beginning teacher;
- Level 5 represents the advanced skills and abilities of a candidate very well qualified and ready to teach. (SCALE, 2013, p. 12)

Cut Scores in Illinois and Georgia. Individual states set their own passing scores for the various content areas and have the authority to increase cut scores over time. During the pilot study of teacher candidate performance on the WL edTPA (Hildebrandt & Swanson, 2014), cut scores were not yet determined for either Illinois or Georgia. Since official implementation began in fall 2015, however, the cut score to pass the WL edTPA in Illinois and Georgia was 29 and 31, respectively, between September 1, 2015 to August 31, 2017. From September 1, 2017 to August 31, 2018, the cut scores were 32 and 33, respectively. Illinois increased the necessary scores for the WL edTPA to 35 in September of 2018 and will again increase it to 37 in September of 2019.

Procedures and Data Analysis. In the aforementioned pilot study, trained evaluators scored the Georgia State University WL edTPA portfolios, and the Illinois State University Office of the Provost provided funding to pay the \$300 fee per portfolio. At Georgia State University, teacher preparation program coordinators locally scored candidates' portfolios following SCALE training. That

Re-examining teacher candidate performance on world language edTPA

is, all Georgia State University teacher candidate portfolios were placed in *LiveText* and scored locally by the second author and another Georgia State University colleague. Inter-rater reliability was strong (Hildebrandt & Swanson, 2014).

In the present study, all WL edTPA portfolios were externally scored by trained evaluators once the portfolios were uploaded to the Pearson site and each participant paid the \$300 fee. Official scores were returned to the participants and to their universities. Then, the anonymous data were entered into an Excel spreadsheet at each university, and the authors combined the two spreadsheets and checked them for accuracy. Afterward, the data were entered into and analyzed using SPSS 25.0. Due to the low number of participants, only measures of central tendency (e.g., frequency counts, means) are reported here.

Findings

The researchers investigated how programmatic changes to their WL teacher education programs impacted edTPA scores since edTPA became consequential for teacher licensure in their states and following the pilot testing phase. They also compared those scores to national averages on the WL edTPA.

With respect to the first research question about the effect of the combined interventions on each of the three tasks (i.e., Planning, Instruction, and Assessment) since 2014, Table 1 shows that the participants' composite score means decreased more than 3.50 points, from 43.12 ($SD = 5.98$) in the previous study (Hildebrandt & Swanson, 2014) to 39.59 ($SD = 5.59$) in the current study. The subscore for the Assessment task, however, had a modest increase between 2014 ($M = 3.04$, $SD = 0.96$) and the time of this study ($M = 3.18$, $SD = 0.56$).

Table 1. Means and standard deviations for the three edTPA tasks from two time periods.

	2014 ($N = 21$)		Fall 2015 - Fall 2017 ($N = 37$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Task 1: Planning	3.64	0.46	3.07	0.47
Task 2: Instruction	3.33	0.56	2.82	0.53
Task 3: Assessment	3.04	0.96	3.18	0.56
Composite edTPA score	43.12	5.98	39.59	5.59

Mean scores on Task 1 (Planning for Teaching and Learning) and Task 2 (Instructing and Engaging Students in Learning) decreased by more than a half a point each between the previous and the current study. Specifically, the Task 1 mean score dropped from 3.64 ($SD = 0.46$) to 3.07 ($SD = 0.47$) and the Task 2 mean score dropped from 3.33 ($SD = 0.56$) to 2.82 ($SD = 0.53$).

NECTFL Review 84

Closer examination of the data from the 13 rubrics across the three tasks (see Table 2) showed that participants in the current study scored highest on Rubric 11 (Providing Feedback to Guide Student Development of Communicative Proficiency in the Target Language, $M = 3.36$, $SD = 0.74$) and Rubric 1 (Planning for Communicative Proficiency in the Target Language, $M = 3.31$, $SD = 0.74$). In the previous study, the highest rubric means were almost all found in Task 1, with the exception of Rubric 5 (Learning Environment, $M = 3.71$, $SD = 0.72$).

Table 2. Means, standard deviations, and frequencies for the 13 WL edTPA rubrics from two time periods.

	2014 ($N = 21$)		2017 ($N = 37$)	
	M	SD	M	SD
Planning				
Planning for Communicative Proficiency in the Target Language (R1)	3.86	0.65	3.31	0.74
Planning to Support Varied Student Learning Needs (R2)	3.57	0.59	3.19	0.73
Using Knowledge of Students to Inform Teaching and Learning (R3)	3.52	0.75	3.20	0.55
Planning Assessments to Monitor and Support Students' Development of Communicative Proficiency (R4)	3.62	0.67	3.04	0.71
Instruction				
Learning Environment (R5)	3.71	0.72	3.14	0.42
Engaging Students (R6)	3.43	0.67	2.91	0.80
Deepening Student Communicative Proficiency in the Target Language (R7)	3.24	0.83	2.89	0.70
Subject-Specific Pedagogy (R8)	3.05	0.92	2.42	0.92
Analyzing Teaching Effectiveness (R9)	3.24	0.70	2.74	0.72
Assessment				
Analysis of Student Communicative Proficiency in the Target Language (R10)	3.38	0.80	3.19	0.84
Providing Feedback to Guide Student Development of Communicative Proficiency in the Target Language (R11)	3.19	1.07	3.36	0.74
Student Use of Feedback (R12)	2.71	1.05	3.06	0.64
Using Assessment to Inform Instruction (R13)	2.90	0.62	3.09	0.66

Re-examining teacher candidate performance on world language edTPA

The lowest mean scores in this study were found on Rubric 9 (Analyzing Teaching Effectiveness, $M = 2.74$, $SD = 0.72$) and Rubric 8 (Subject-Specific Pedagogy, $M = 2.42$, $SD = 0.92$), both of which are part of Task 2, Instructing and Engaging Students in Learning. In the previous study, Rubrics 12 (Student Use of Feedback, $M = 2.71$, $SD = 1.05$) and 13 (Using Assessment to Inform Instruction, $M = 2.90$, $SD = 0.62$), both from the Assessment task, had the lowest means.

Mean scores for Rubrics 12 and 13 increased 0.35 and 0.19 respectively from the 2014 study to the present study, along with the mean score for Rubric 11 (Providing Feedback to Guide Student Development of Communicative Proficiency in the Target Language), which increased by 0.17. All increases in mean scores were found in the Assessment task, while mean scores for all rubrics in the Planning and Instruction tasks decreased. Mean scores for the remaining 10 rubrics decreased, with differences ranging from 0.19 in Rubric 10 (Analysis of Student Communicative Proficiency in the Target Language) to 0.63 in Rubric 8 (Subject-Specific Pedagogy), with five of the 10 rubrics decreasing by more than half a point.

Turning to the second research question concerning how participants' scores compared to national WL edTPA means, data analysis revealed that that national mean for the WL edTPA dropped 4.06 points from 2014 to 2016 (see Table 3). In similar fashion, the total mean score for the two focus WL teacher preparation programs decreased 3.53 points from 2014 to 2017. However, mean composite scores for the two programs were consistently higher than national averages.

Table 3. Means and standard deviations for national and programmatic WL edTPA composite scores.

	<i>M</i>	<i>SD</i>
2014 National WL edTPA Total Score ($N = 416$)	40.00	7.73
2015 National WL edTPA Total Score ($N = 572$)	37.24	7.30
2016 National WL edTPA Total Score ($N = 655$)	35.94	6.47
2014 Program WL edTPA Total Score ($N = 21$)	43.12	5.98
2015-2017 Program WL edTPA Total Score ($N = 37$)	39.59	5.59

Discussion

At the time of the 2014 pilot study, colleges and universities in 34 states and the District of Columbia were in various stages of edTPA implementation (Hildebrandt & Swanson, 2014). Currently, the number of states with institutions of higher education that use edTPA has grown to 41, with 19 states using the assessment for

licensure or program accreditation decisions (AACTE, 2018b). Such growth in use increases the field's urgency to examine the assessment's ability to fairly and accurately measure new teacher readiness. Its high-stakes nature, cost, and influence on teacher education curricula necessitate a critical evaluation of score patterns. The present study compared two WL teacher education programs' edTPA outcomes across time and circumstances. The first study was conducted before edTPA became consequential in Illinois and Georgia, and the second was carried out after it became consequential for individual teacher licensure in both states and graduation from teacher education programs at Illinois State University. The researchers also compared trends within the two programs to those within national WL edTPA data to determine larger patterns of teacher candidate performance.

Results from the initial pilot study indicated that participants scored the highest on Tasks 1 and 2, while the Assessment task proved to be the most challenging (Hildebrandt & Swanson, 2014). The researchers theorized that these results were found because teacher candidates have many opportunities to plan lessons and enact them; however, these teacher candidates rarely have opportunities to gather and analyze student data that result from their instruction. Even though the participants' scores were above the cut scores necessary for Illinois and Georgia teacher licensure at the time, the researchers redesigned the two programs curricula in an attempt to address the lower Task 3 (i.e., Assessing Student Learning) scores. While there were concerns that an assessment-focused curricular redesign could negatively influence participants' scores on Tasks 1 and 2, the authors, who are the directors of the two programs, agreed that placing a heavier focus on assessment was warranted. Due to programmatic requirements (e.g., maximum number of credit hours in the programs), it was not feasible to add a new course on assessment. In similar fashion, changing the curricula during earlier field experiences (e.g., practica) to focus on gathering and analyzing student data proved challenging. In both programs, completers earn a multiple-level teacher certification (i.e., P-12 in Georgia, K-12 in Illinois), and they have field experiences at all levels of instruction. Analysis of the 2015-2017 WL edTPA data (SCALE, 2015, 2016a, 2017a) show that the redesign was indeed successful at raising the Assessment task subscore, albeit modestly, with the enacted curricular modifications having a positive impact on three of the four Assessment task rubrics.

Specifically, yet briefly, the authors focused curricular changes based on principles of backwards design, writing active and measurable learning outcomes (Swanson & Hildebrandt, 2017), ACTFL's Core Practices, and developing an integrated performance assessment for use in teacher candidates' WL edTPA portfolios. During practicum classes, program faculty placed even more emphasis on teacher candidates starting with the end in mind (Wiggins & McTighe, 2005) by beginning with the assessment task when constructing their portfolios (Hildebrandt & Swanson, 2016). Additional emphasis was placed on the value of writing measurable learning outcomes during methods classes (Swanson & Hildebrandt, 2017) while integrating strategies to build a classroom discourse community (Glisan & Donato, 2017). Such changes clearly had a positive effect on improving Task 3 scores.

Re-examining teacher candidate performance on world language edTPA

Unfortunately, the focus on assessment practices had a negative impact on scores for all Task 1 and 2 rubrics, as compared to the 2014 study, because of curricular and time constraints to move teacher candidates through the programs in a timely manner. Initially, the researchers were hesitant to share results that answered the first research question concerning the impact of interventions put in place. After all, the programs had additional time to emphasize concepts addressed in the WL edTPA and, yet, the teacher candidates were not performing as well, both overall (as shown by the drop in composite score means) and on 10 of the 13 rubrics. Upon answering the second research question comparing the two programs' data to national data, however, the researchers realized that decreases in scores were a national issue. Mean scores for WL edTPA composite scores and almost all 13 rubrics have steadily fallen across the nation (SCALE, 2015, 2016a, 2017a) since the initial study. As the number of national WL edTPA submissions increased from 420 in 2014 (SCALE, 2015) to 587 in 2015 (SCALE, 2016a) to 672 in 2016 (SCALE, 2017a), the mean composite scores dropped from 40.00 to 37.20 to 35.90 points, respectively. Additionally, national means on all but three of the 13 rubrics declined each year; those three rubrics (Rubrics 10, 11, and 12) showed an initial decline from 2014 to 2015 and in 2016 either rebounded to the 2014 level (Rubric 12) or plateaued (Rubrics 10 and 11). Such a general downward trend may be accounted for, in part, by the increase in submissions and additional educator preparation programs inexperienced with edTPA. Nevertheless, the fact that data from the two programs showed a similar pattern for 10 of the 13 rubrics calls such an explanation into question. The substantial curricular changes made to our programs in order to better prepare our teacher candidates, the increased familiarity with the assessment among teacher candidates and program directors, and the licensure mandate in our states did not seem to have made much of a positive impact on composite scores. The positive impacts of interventions on the three assessment rubrics with higher means were modest at best.

An additional concern, both in the two focus programs and across the nation, is the decrease in mean scores for Rubric 8 (Subject-Specific Pedagogy). This rubric, which is part of the Instruction task, addresses four of the five goal areas from the *World-Readiness Standards* (The National Standards Collaborative Board, 2015)—Communication, Cultures, Comparisons, and Communities. It specifically demands that teacher candidates address all three Ps (i.e., products, practices, and perspectives) from the Cultures goal area in order to earn a score of 2 of on the rubric. In order to earn a higher score, teacher candidates must connect those three Ps to the students' existing knowledge and experiences and encourage the students to use the target language beyond the classroom setting. These expectations may be appropriate if examining the entire learning segment of three to five lessons; they are not appropriate, however, for examining the primary sources of evidence for Rubric 8—one 15-minute, unedited video and one prompt of the written instructional commentary. Demonstrating the impact of such an unrealistic expectation, national data from 2014 to 2016 show that of the 13 rubrics, Rubric 8 consistently has the lowest mean (SCALE, 2015, 2016a, 2017a). The same is true for this study's data set, with the mean for Rubric 8 more

than 0.30 points lower than the next lowest mean (i.e., Rubric 9, also within the Instruction task). Such a finding is particularly vexing since an edTPA task force from ACTFL approached members of SCALE several years ago, presenting issues related to the assessment including Rubric 8. At the time, the task force was assured that the concerns expressed would be taken into consideration when revising the subsequent WL edTPA handbook. However, minimal changes or revisions were made to that handbook, including changes to Rubric 8. In the subsequent version of the WL edTPA handbook's glossary, more information was provided about the Communication goal's three modes and the Cultures goal's perspectives, products, and practices (The National Standards Collaborative Board, 2015), but Rubric 8 remained problematic in spite of the field's efforts to revise it. The existing unfeasible expectations of Rubric 8 not only call into question the reliability and validity of the WL edTPA; they threaten the WL teacher supply across the nation as teacher candidates may miss required cut scores, in part, because of their low score on the rubric.

An overall decrease in total scores at both the programmatic and national levels is evident and alarming. From 2014 through 2016, the national WL edTPA scores dropped over four points. Each year mean scores on nearly every rubric decreased (SCALE, 2015, 2016a, 2017a). A similar trend is found with respect to the program scores for the current study, with composite score means dropping over three points from 2014 to 2017. These findings are particularly concerning given that scores required for licensure in Illinois and graduation from Illinois State University will increase to 37 in the fall of 2019 (AACTE, 2018b). With a composite mean score of 39.59 across the two programs, adverse consequences may arise for a teacher education program with higher numbers of teacher candidates who cannot meet cut scores on the first attempt, not to mention the financial implications for teacher candidates themselves.

An overall decrease in total scores at both the programmatic and national levels is evident and alarming.

At present, all 50 states are reporting shortages of world language teachers (U.S. Department of Education, 2016). An increase in cut scores may also exacerbate the already severe WL teacher shortage in both states as well as the nation as a whole (Swanson & Mason, 2018). One of the major issues is teacher attrition (Swanson, 2012). Perda (2013) found that more than 42% of new teachers leave the profession within five years of entry, which supports longitudinal data showing a steady increase in beginning teacher attrition over the past 2 decades (Ingersoll & Merrill, 2013). A result of such attrition is a lack of credentialed teachers. At present, public schools are having difficulties finding certified teachers, especially in world languages (Hanford, 2017; Koerting, 2017), and districts are hiring non-certified teachers on emergency credentials. These non-certified teachers, however, will not have completed an edTPA portfolio to demonstrate their teaching abilities, frustrating the original intent of laws focused on teacher performance assessments. The pipeline feeding schools is not much better. Long (2016) reported that “enrollment in teacher education programs is down significantly—falling 35% nationwide in the last five years” (p. 1).

Re-examining teacher candidate performance on world language edTPA

Also accompanying higher edTPA cut scores and lower average performance are threats to world language teacher education programs that face closure due to low numbers, if not for poor performance by teacher candidates on standardized assessments. When teacher education requirements are as demanding as they have ever been, extraneous variables outside the control of individual teacher candidates and teacher education programs may make entering the teaching profession unnecessarily difficult for language teacher candidates at a time when they are severely needed in the field.

With these findings, the authors suggest that state boards of education and legislatures consider adjusting cut-scores that have ramped up over the course of several years to more realistic levels. After all, overall means have dropped in national data, despite an increased familiarity with the assessment among teacher education programs and teacher candidates. New York serves as a precedent for this practice. Originally, New York set 35 as a minimum cut score on the WL edTPA through 2017, but then decreased it to 32 through 2019, gradually ramping up again to 34 after 2022 (SCALE, 2018). After all, cut scores do not have to remain stagnant once set and may be adjusted as necessary to regulate supply of teachers into a state's schools.

With these findings, the authors suggest that state boards of education and legislatures consider adjusting cut-scores that have ramped up over the course of several years to more realistic levels.

Given the findings from this study and the overall national longitudinal trends regarding the WL edTPA, one final implication remains—there remain so are many unanswered questions regarding this high-stakes assessment. First, why do WL edTPA scores continue to decrease both locally and nationally? Is the decrease attributable to the quality of the individual creating the portfolio or to other unknown variables related to teacher preparation programs? This set of questions related to preservice teacher quality then cascade to the next set concerning the quality of portfolio reviewers. Who are the people serving as external reviewers? How long has each served as an external reviewer? How many external reviewers are there and is there a shortage of them? What qualifies them to serve as reviewers of beginning teachers beyond what is stated on their website (Pearson Technology, 2018)? Such questions regarding the external call into question the accountability applied to teacher education programs and individual teacher candidates. If variables beyond their control influence scores, how can individuals and programs be held accountable? Finally, with decreases on all three tasks of the WL edTPA, it is evident that the lower scores are not just a local issue, but rather a national issue, especially in terms of Rubric 8 (Subject-Specific Pedagogy).

Limitations

Given these findings and implications of this research, the authors note the study's limitations. First, data are from only two programs and the number of participants is too low to be able to carry out analyses using inferential statistics. Small scale studies' findings must be interpreted with caution. Further, some of the original study's data, which were used as a comparison for the more recent data,

NECTFL Review 84

were generated by program faculty who locally scored the WL edTPA portfolios, not official Pearson scorers. Although those faculty completed the WL edTPA scoring training, the scores cannot be considered official.

Future Research

Nevertheless, the researchers call for more research on the WL edTPA. It would be insightful to continue to examine the aforementioned trends from a longitudinal perspective, both from a program and national perspective, as well as from a qualitative perspective. It would also be of interest to investigate whether these trends exist in other content areas. Additionally, research from other WL teacher education programs discussing initial findings and any programmatic changes to address teacher candidate success would be helpful. Such research would provide opportunities for collaboration among smaller programs to address issues surrounding the WL edTPA. Since its inception on the national stage, edTPA has grown immensely and is now in over 800 Educator Preparation Programs in 41 states and the District of Columbia (AACTE, 2018a). edTPA is a high-stakes performance assessment that concerns teacher candidates and teacher education faculty alike, and its pervasive use for decisions around graduation and teacher licensure is alarming given its problematic nature (Hildebrandt & Swanson, 2016). Considering that program faculty express concern and worry that WL teacher education programs will be evaluated by their teacher candidates' ability to pass edTPA, more research is clearly warranted. Teacher effectiveness is crucial, but effective assessment of teacher quality is equally important.

References

- American Association for Colleges of Teacher Education. (2018a). *Participation map*. Retrieved December 20, 2018 from, <http://edtpa.aacte.org/state-policy>
- American Association for Colleges of Teacher Education. (2018b, June). State edTPA policy overview. Retrieved December 20, 2018 from, https://secure.aacte.org/apps/rl/res_get.php?fid=1014&ref=edtpa
- American Council on the Teaching of Foreign Languages. (2010). Use of the target language in the classroom. Retrieved December 20, 2018 from, <https://www.actfl.org/news/position-statements/use-the-target-language-the-classroom>
- American Council on the Teaching of Foreign Languages. (2013). *ACTFL/CAEP program standards for the preparation of foreign language teachers*. Retrieved December 20, 2018 from, https://www.actfl.org/sites/default/files/CAEP/ACTFLCAEPStandards2013_v2015.pdf
- American Council on the Teaching of Foreign Languages. (2015). Building your core. Retrieved December 20, 2018 from, <https://www.actfl.org/sites/default/files/pd/presentations/2016/Building%20Your%20Core%20-%20Effective%20Practices.pdf>
- Au, W. (2013). What's a nice test like you doing in a place like this? The edTPA and corporate education "reform." *Rethinking Schools*, 27(4), 22-27.

Re-examining teacher candidate performance on world language edTPA

- Behney, J. (2016). Perceived influence of cooperating teachers on edTPA performance. *Foreign Language Annals*, 49(2), 271-286.
- Behney, J. (2018). Performance on the assessment and instruction tasks of the World Language edTPA and avenues for further research. In P. Swanson & S. A. Hildebrandt (Eds.), *Researching edTPA promises and problems* (pp. 187-200). Charlotte, NC: Information Age Publishing.
- Cochran-Smith, M., Piazza, P., & Power, C. (2013). The politics of accountability: Assessing teacher education in the United States. *The Educational Forum*, 77(1), 6-27.
- Dover, A. G., Schultz, B. D., Smith, K., & Duggan, T. J. (2015). Who's preparing our candidates? edTPA, localized knowledge and the outsourcing of teacher evaluation. *Teachers College Record*, March 30. Retrieved December 20, 2018 from, <http://www.tcrecord.org/content.asp?contentid=17914>
- Georgia State University. (2019). About. Retrieved July 16, 2019 from, <https://www.gsu.edu/about/>
- Glisan, E. W., & Donato, R. (2017). *Enacting the work of language instruction: High-leverage teaching practices*. Alexandria, VA: ACTFL.
- Gronlund, N. E. (2004). *Writing instructional objectives for teaching and assessment*. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Hanford, E. (2017). Schools in poor, rural districts are the hardest hit by nation's growing teacher shortage. APM Reports. Retrieved December 20, 2018 from, <https://www.apmreports.org/story/2017/08/28/rural-schools-teachershortage>
- Hildebrandt, S. A., Hlas, A. C., & Conroy, K. (2013). Comparing nine world language teacher preparation programs within and across three states. *NECTFL Review*, 72, 15-44. Retrieved from <https://www.nectfl.org/?ddownload=1578>
- Hildebrandt, S. A., & Swanson, P. (2014). World language teacher candidate performance on edTPA: An exploratory study. *Foreign Language Annals*, 47, 576-591.
- Hildebrandt, S., & Swanson, P. (2016). *Understanding the world language edTPA: Research-based policy and practice*. Charlotte, NC: Information Age Publishing.
- Illinois State University. (2018). University facts. Retrieved December 20, 2018 from, https://prpa.illinoisstate.edu/data_center/university/Fall%202018%20Fact%20Pamphlet%20Final.pdf
- Illinois State University. (2019). College of Education: About. Retrieved July 16, 2019 from, <https://education.illinoisstate.edu/about/>
- Ingersoll, R., & Merrill, E. (2013). Seven trends: The transformation of the teaching force. CPRE Report. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.
- Jourdain, S. (2018). Native / heritage Spanish speakers and edTPA. In P. Swanson & S. A. Hildebrandt (Eds.), *Researching edTPA promises and problems: Perspectives from English as an additional language, English language arts, and world language teacher education* (pp. 67-86). Charlotte, NC: Information Age Publishing.
- Koerting, K. (2017). Schools confront shortage of world language teachers. *News Times*. Retrieved December 20, 2018 from, <http://www.newstimes.com/local/article/Schools-confront-shortage-of-world-language-10996278.php>

NECTFL Review 84

- Long, C. (2016, September 19). Report: Teacher shortage crisis can be averted by keeping educators in the profession. NEA Today. Retrieved December 20, 2018 from, <http://neatoday.org/2016/09/19/teacher-shortage-crisis/>
- Madeloni, B., & Gorlewski, J. (2013). Wrong answer to the wrong question: Why we need critical teacher education, not standardization. *Rethinking Schools*, 27, 4. Retrieved December 20, 2018 from, http://www.rethinkingschools.org/archive/27_04/27_04_madeloni-gorlewski.shtml
- National Standards for Foreign Language Education Project. (2006). *Standards for foreign language learning in the 21st century* (3rd ed.). Lawrence, KS: Allen Press. Author.
- Pearson Technology. (2018). *edTPA scorer qualifications*. Retrieved December 20, 2018 from, <http://scoreedtpa.pearson.com/become-an-edtpa-scorer/edtpa-scorer-qualifications.html>
- Perda, D. (2013). Transitions into and out of teaching: A longitudinal analysis of early career teacher turnover (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia.
- Russell, V., & Davidson Devall, K. F. (2016). An examination of the edTPA portfolio assessment and other measures of teacher preparation and readiness. *Foreign Language Annals*, 49(4), 479-501.
- Stanford Center for Assessment, Learning, and Equity. (2013). *edTPA world language assessment handbook*. Stanford, CA: Stanford University.
- Stanford Center for Assessment, Learning, and Equity. (2015). *Educative assessment & meaningful support: 2014 edTPA administrative report*. Retrieved December 20, 2018 from, https://secure.aacte.org/apps/rl/res_get.php?fid=2183&ref=edtpa
- Stanford Center for Assessment, Learning, and Equity. (2016a). *Educative assessment & meaningful support: 2015 edTPA administrative report*. Retrieved December 20, 2018 from, https://secure.aacte.org/apps/rl/res_get.php?fid=3013&ref=rl
- Stanford Center for Assessment, Learning, and Equity. (2016b). By the numbers: Educative assessment & meaningful support: 2016 edTPA annual administrative report Retrieved December 20, 2018 from, https://secure.aacte.org/apps/rl/res_get.php?fid=3638&ref=edtpa
- Stanford Center for Assessment, Learning, and Equity. (2017a). *Educative assessment & meaningful support: 2016 edTPA administrative report*. Retrieved December 20, 2018 from, https://secure.aacte.org/apps/rl/res_get.php?fid=3621&ref=rl
- Stanford Center for Assessment, Learning, and Equity. (2017b). *edTPA world language assessment handbook*. Stanford, CA: Stanford University.
- Stanford Center for Assessment, Learning, and Equity. (2018, May). *edTPA for New York state*. Retrieved December 20, 2018 from, https://www.edtpa.com/PageView.aspx?f=GEN_NewYork.html
- Swanson, P. (2012). The congruence of vocational interests and the workplace environment: Reducing the language teacher shortage. *Language Teaching Research*. 16(4), 519-537.
- Swanson, P., & Goulette, E. (2018). The criticality of the integrated performance assessment in the world language edTPA. In P. Swanson & S. A. Hildebrandt (Eds.), *Researching edTPA promises and problems: Perspectives from English as an additional language, English language arts, and world language teacher education* (pp. 201-220). Charlotte, NC: Information Age Publishing.

Re-examining teacher candidate performance on world language edTPA

- Swanson, P., & Hildebrandt, S. (2017). Communicative learning outcomes and world language edTPA: Characteristics of high-scoring portfolios. *Hispania*, 100, 331-347.
- Swanson, P., & Mason, S. (2018). The world language teacher shortage: Taking a new direction. *Foreign Language Annals*, 51(1), 251-262.
- The National Standards Collaborative Board. (2015). *World-readiness standards for learning languages* (4th ed.). Alexandria, VA: Author.
- Troyan, F. J., & Kaplan, C. S. (2015). The functions of reflection in high-stakes assessment of world language teacher candidates. *Foreign Language Annals*, 48(3), 372-393.
- U.S. Department of Education. (2016). Teacher shortage areas nationwide listing. 1991 through 2017-2018. Retrieved December 20, 2018 from, <https://www2.ed.gov/about/offices/list/ope/pol/ateachershortageareasreport2017-18.pdf>
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Alexandria, VA: ASCD.