Article

Educational Policy
2018, Vol. 32(5) 728–750

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0895904816673741
journals.sagepub.com/home/epx
SAGE

A Rational and Manageable Value-Added Model for Teacher Preparation Programs

Michael P. Brady[1], Katie M. Miller[1], Jazarae McCormick[1], and Lawrence A. Heiser[1]

Abstract

Educators struggle with "value-added" teacher evaluation models based on high-stakes student assessments. Despite validity and reliability threats, these models evaluate university-based teacher preparation programs (TPPs), and play a role in state and professional accreditation. This study reports a more rational value-added evaluation model linking student performance to teacher candidates' lessons during Practicum and Student Teaching. Results indicate that K-12 students showed learning gains on these lessons, with mixed findings on comparisons of part-time to full-time internships, academic and functional lessons, and candidates' grade point averages (GPAs). Results indicated that teacher candidates' lessons are a viable value-added model (VAM) alternative for TPPs.

Keywords

value-added models, VAM, validity and reliability threats, teacher preparation programs

[1]Florida Atlantic University, Boca Raton, FL, USA

Corresponding Author:
Michael P. Brady, Department of Exceptional Student Education, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA.
Email: mbrady@fau.edu

A wrinkle in the complex world of university-based teacher preparation includes accountability mandates that require teacher preparation programs (TPPs) to adopt teacher evaluation models based on K-12 student achievement. With little grounding in evidence-based practice, these "value-added" mandates define TPPs as effective if their teacher candidates and graduates produce learning gains in K-12 students. These value-added models (VAMs) of accountability require that TPPs

- deliver measurably high-quality instruction to teacher candidates,
- show that teacher candidates learn to perform as teachers as a result of this instruction,
- show that teacher candidates and completers then deliver high-quality instruction to K-12 students as an outcome of participating in the program, and
- demonstrate that the K-12 students of the candidates and completers produce better academic achievement gains as a result of their teachers' instruction.

Most value-added assessment models use statewide student achievement assessments as the data for teacher, and now TPP, evaluations. These standardized assessments frequently are limited to a few grade levels and select content areas, so teachers outside of these assessment targets are not well represented in the data that are used for their evaluations (Steinbrecher, Selig, Cosbey, & Thorstensen, 2014). Also, using student assessments as the data for making individual teacher and wider scale programmatic decisions is well beyond both the intended use of the data and the norming efforts of the tests. This presents a direct challenge to the validity of the assessments (Ballou & Springer, 2015; Hill, Kapitula, & Umland, 2011; Lincove, Osborne, Dillon, & Mills, 2014). Researchers have reported serious issues involving the reliability of VAMs based on student assessment (Amrein-Beardsley, 2008; Berliner, 2014). Papay (2011), for example, found that student standardized test results are unreliable when predicting teacher performance and across multiple assessments of students' performance. Others have reported unstable measurement characteristics over time (McCaffrey, Sass, Lockwood, & Mihaly, 2009) and across subscales of the same assessments (Lockwood et al., 2007), with even subtle changes in statistical procedures moving "high performing" teachers and TPPs to "low performing" status (Lincove et al., 2014).

Additional challenges to using standardized assessments to evaluate teachers and TPPs involve the inability of these tests to adequately assess learning in English Language Learners, students with disabilities, and nonfluent readers who demonstrate complex subject matter mastery when evaluated with authentic learning assessments (Holdheide, Goe, Croft, & Reschly, 2010; Steinbrecher et al., 2014). Because these student subgroups frequently perform poorly on standardized assessments, policy makers erroneously ascribe poor student performance to ineffective teaching (Berliner, 2014; Buzick & Laitusis, 2010; Jones, Buzick, & Turkan, 2013). These K-12 students may perform academic assignments with mastery but still remain below an established floor on a standardized test (Buzick & Laitusis, 2010). The unintended consequences of this "system error" are dramatic as teachers learn to shop for the "money kids" (Berliner, 2013) whose test performance will result in improvements in their profiles as effective teachers. Berliner found that these teachers learn to resist efforts to teach low-achieving students. Other unfortunate consequences arise when teachers shift their instructional energies from teaching activities that do not directly prepare students to take standardized tests (Morganstein & Wasserstein, 2014), or when TPPs are unable to place teacher candidates in schools due to teacher concerns that student teachers might reduce their attention to preparing K-12 students for high-stakes assessments, thus risking these teachers' own evaluations (Brady, Duffy, Hazelkorn, & Bucholz, 2014; Moore Johnson, 2015).

A third set of challenges has been identified involving the difficulty in isolating the effects of individual teachers on student learning. Current VAM systems track K-12 students' learning gains across multiple teachers, within and across multiple school years. Isolating the effects of individual teachers is a fundamental assumption of VAM accountability efforts, yet numerous practices interfere with this assumption (Berliner, 2014; Floden, 2012; McCaffrey et al., 2009). As they move across three grade levels in 3 years, most K-12 students have a minimum of three different teachers who contribute to their learning. Students who receive pull-out or supplemental instruction (e.g., English Language Learners, students with disabilities, others who receive remedial instruction) typically have even more teachers who contribute to their learning. In schools with "shared teacher" arrangements, co-teachers, and block scheduling, students could have six or more teachers during that same period. If schools accept teacher candidates during their Practicum, Internship, or Student Teaching experiences, any number of teachers might contribute to students' learning. Across a 3-year time span, K-12 students might be taught by more than a dozen teachers, for distinctly different lengths of time, all of whom make differential contributions to their learning.

In spite of these challenges, TPPs continue to explore ways to evaluate teacher candidates (American Educational Research Association [AERA], 2015). To minimize the validity and reliability flaws inherent in current VAMs, Brady, Heiser, McCormick, and Forgan (2016) proposed minimum standards of evidence-based practice for VAM assessments. These standards include

1. Links between teacher candidates' performance and K-12 student learning gains should be explicit, and limited to the actual candidates and their students;
2. Student learning gains should be based on curriculum targets and content actually delivered to them by the teacher candidates;
3. Student learning gains should inform teacher candidates of their effectiveness in teaching by providing feedback that acknowledges or improves their teaching (i.e., information on student performance actually is useful for further teacher development); and
4. Student learning gains should be based on multiple, authentic work samples rather than standardized achievement tests.

To date, few alternatives to standardized test scores have been adopted as VAMs (AERA, 2015; Darling-Hammond, 2015). Elliott, Roach, and Kurz (2014) advocated teacher portfolios as a VAM, many of which include observations and teacher self-report. One portfolio assessment is the edTPA, which serves as a culminating evaluation for teacher candidates (American Association of Colleges for Teacher Education [AACTE], n.d.; Pecheone, Shear, Whittaker, & Darling-Hammond, 2013). These portfolio systems are positive alternatives to VAM based on standardized assessments, but are limited by their expense, resource, and time implications. Jones and Brownell (2014) advocated a different approach to VAM based on teachers' classroom observations, and Goldring et al. (2015) demonstrated that when principals have access to high-quality observation systems, they use these data rather than standardized test scores for decision making. Although observation methods are well known to teacher educators, different observation systems frequently yield vastly different profiles of the same teachers (Hamilton & Brady, 1991; Swank, Taylor, Brady, & Freiberg, 1989). For example, whole-class observations more easily identified teachers who were "on task" (i.e., obviously delivering a lesson, regardless of students' engagement), while observations of teachers with individual students yielded information on the impact of instructional variables such as academic questions, corrections, and reinforcement. When comparing observations of teachers interacting with specific high- and low-achieving students versus observations of teachers with the entire class, only 37% of the teachers were identified as effective by both observation systems (Swank et al., 1989).

A recent exploration of an alternative VAM was conducted by Brady et al. (2016) with undergraduate and graduate students enrolled in a university TPP. This study reported learning gains on student performance on actual lessons delivered by the teacher candidates who were being evaluated, based on requirements in the teacher preparation curriculum. The findings showed that for three cohorts of candidates, K-12 students showed gains of at least 35% between pretest and posttest measures in the lessons delivered by the candidates, with only a small number of candidates' students showing less than a 10% learning gain. Across the cohorts, between 73% and 96% of the K-12 students met the learning objectives for their lessons. In addition, the investigators discovered that most candidates needed university supervisor support to report their K-12 students' progress as gain scores and to revise their lessons based on this performance. These findings were used to generate improvements to the TPP, including changes in the instructional methods course for subsequent teacher candidates.

In the remainder of this article, we present the results of an investigation exploring an alternative VAM based on teacher candidates' delivery of structured lessons during their Practicum and Student Teaching experiences. In this study, we sought to replicate the alternative VAM reported by Brady et al. (2016), and to extend those results by exploring variables not considered in that research. The purpose of this study was first to examine whether the candidates' instruction would result in measurable learning gains in their K-12 students. We also investigated whether the students actually met their learning objectives. Finally, we sought to discover an initial set of variables that might predict whether teacher candidates' instruction might yield student learning gains. We posed five research questions to guide our investigation.

Research Questions

Research Question 1: As a result of participating in instructional activities in a Learning Sequence with the teacher candidates, do K-12 students show learning gains as measured by their pretest and posttest scores on their learning objectives?

Research Question 2: As a result of participating in these instructional activities in a Learning Sequence with the teacher candidates, do K-12 students' learning gains result in meeting their specific objectives?

Research Question 3: Is there a difference in K-12 students' pre–post learning gains as teacher candidates progress from Practicum to Student Teaching? Is there a difference in the learning objectives met by K-12 students as the candidates progress from Practicum to Student Teaching?

Research Question 4: Does the type of lesson delivered by teacher candidates (i.e., academic vs. functional daily living skills) predict whether their K-12 students are likely to show learning gain scores or meet the specific objectives of the Learning Sequence?

Research Question 5: Is there a difference in K-12 students' learning gains, or are their learning objectives met, when taught by candidates with different grade point averages (GPAs) in their teacher preparation coursework?

Method

Participants and Settings

Data were collected from a total of 30 undergraduate teacher candidates enrolled in a bachelor's degree program in special education in a state university in south Florida. These candidates represented the population of students formally enrolled and matriculating through the program during their fall instructional methods Practicum, and their spring Student Teaching in local schools. During the fall Practicum, candidates participated in 90 hr of fieldwork for 14 weeks; during Student Teaching, candidates participated in a full-time daily school placement (37.5 hr per week) for 16 weeks. Fall Practicum accompanied a didactic class that focused on teaching methods for students with mild disabilities, and a classroom management class. Numerous classroom assignments required the teacher candidates to practice or implement what they learned in a local school placement. Throughout the spring, candidates participated in two condensed courses during their Student Teaching. These courses were delivered as workshops targeting Individualized Education Program (IEP) development and collaboration strategies with teachers and families.

Teacher candidates taught students in a variety of K-12 classrooms. K-12 students ranged in age from 7 to 20.5 years (M = 13.5; SD = 3.89) in the Practicum semester, and from 5 to 19 years (M = 10.65; SD = 4.13) during Student Teaching. Candidates had different teaching assignments during the two semesters so that each candidate had both elementary and secondary students during the year. During their Practicum, teacher candidates taught 131 K-12 students (class size M = 4.68; SD = 4.91) across 26 different schools; during Student Teaching, the candidates taught 122 K-12 students (class size M = 4.69; SD = 1.38) in 27 schools. School placements in Practicum included 20 self-contained or special education resource classes, and six general education classes for students with and without disabilities. Student Teaching placements included 19 self-contained or special education resource classes and seven general education classes.

Candidates' Use of the Learning Sequence

Teacher candidates link the content of their didactic instructional methods courses to field experiences by implementing a "Learning Sequence" assignment during Practicum and Student Teaching. Instructional content of the Learning Sequence is based on curriculum standards and the IEPs of K-12 students with disabilities. Candidates assess student performance on their lessons to establish whether or not students make learning gains, and whether they meet the objectives of their lessons. University supervisors and the cooperating classroom teachers supervise the candidates while they deliver these lessons, note the impact of candidates' instruction on students, and provide feedback for supplemental or remedial instruction as needed. If needed, candidates revise and re-teach lessons in the Learning Sequence, add additional lessons, or continue with their original instruction based on student performance.

The Learning Sequence incorporates principles of explicit instruction; after identifying students' instructional needs, candidates establish a specific instructional objective. This objective is task analyzed, and the lesson is sequenced into a minimum of eight explicit lessons. Learning Sequences are diverse in their content. Learning Sequences for some students with moderate to severe disabilities target functional skills (e.g., preparing meals, completing a job application); for students with mild to moderate disabilities, the Learning Sequence typically targets standard academic curriculum (e.g., numeric literacy, reading comprehension). Candidates' lessons require frequent assessment to establish the impact of their instruction on student learning. Prior to lesson delivery, candidates administer an initial pretest. Next, a variety of knowledge and skill assessments accompany instruction, consistent with standard curriculum-based measurement practices. Finally, posttests established overall learning gains on the Learning Sequence lesson(s). Of particular note is whether the K-12 students met the mastery criterion for their lessons.

Behavioral Measures

Two dependent measures were collected. To address whether teacher candidates were effective in teaching the K-12 students, the first measure involved students' change scores on lessons delivered by candidates. Each candidate delivered a number of lessons from a Learning Sequence they designed for students in their field placements. Prior to delivering a lesson, candidates administered a pretest on the objective of the lesson. Following the pretest, candidates delivered 8 to 10 weeks of instruction designed to teach this objective which, in turn, was followed by a posttest. Because candidates were the only individuals who delivered the lessons, student learning gains between the pretest and posttest were attributed to candidates' instructional efforts.

The second dependent measure established whether the students met the learning objective that the candidates taught. For example, if a candidate taught a geography lesson in which students were required to differentiate longitude versus latitude, the criteria for meeting the learning objective would be to differentiate between the two indicators. If seven of eight students met the criteria set for lesson mastery, then 87.5% of those K-12 students were counted as meeting the criteria.

In addition to these outcome measures, two other measures were collected to (a) differentiate the types of lessons delivered by the teacher candidates and (b) establish whether candidates' *GPA within the major* might predict their success is promoting learning in K-12 students. For lesson type, we simply compared the nature of the lessons delivered to the standardized curriculum delivered in the state. Where candidates' lessons matched the academic lessons proscribed in the curriculum, we coded them as "academic." Lessons that did not match the academic curriculum and targeted daily living or social skill learning were coded as "functional." Finally, for the measures for Research Question 5, we reviewed the candidates' GPA in the teacher preparation courses in the special education major. The GPA included only courses that were assigned a letter grade, and included 30 credits distributed among 11 didactic courses. Although candidates enrolled in other teacher preparation classes outside of their major (e.g., TESOL and Foundations courses), these courses were not included in the analysis of the *GPA in the major*.

Data Collection Procedures

Data were collected by asking candidates to complete a 21-item questionnaire used by the program to monitor their Practicum and Student Teaching activities. (This questionnaire was developed by program faculty who taught the undergraduate teacher education classes, with input from staff who supervised candidates' field placements.) First, the questionnaire asked candidates to describe the classroom in which they were teaching, the number and types of students in the class, and the content of the curriculum. Second, candidates were asked to identify a lesson within the Learning Sequence that they planned to deliver to students in the class. Information on the lesson included the objective to be taught, criteria for mastery, and the scope and sequence of instruction that would lead to mastery. Third, candidates were asked to deliver, and then record the pretest scores for all students who received their instruction. Fourth, as the semester drew to a close, candidates were asked to deliver and record posttest scores for this instruction. Finally, candidates were asked to supply examples of instructional materials used in the lesson, and work samples or other evidence of student performance. Two procedures were used to assure the fidelity of the information provided by candidates. First, the cooperating classroom teachers worked with the candidates to assure that all information on the school, placements, and K-12 students was accurate; second, each candidate's university supervisor observed the implementation of the lesson and confirmed accuracy of the lesson content and student performance results. Candidates uploaded the questionnaire and attachments to an electronic collection site; two of the investigators reviewed the information, noted whether the lesson was academic or functional, and sought clarification if needed. The data collection procedure for candidates' GPAs involved collecting the information from electronic transcripts.

Data Analysis

All data from the questionnaires and university databases were loaded to an Excel file, and then to SPSS (version 22). To address the first research question (whether K-12 students showed learning gains), we compared differences on their pretest and posttest scores on the Learning Sequence delivered by the candidates, and whether these change score differences were statistically significant. This analysis was conducted separately for Practicum and Student Teaching semesters so we could ascertain the results of each. To analyze these differences, we conducted a dependent-sample t test. Because not all of the candidates in Practicum continued directly into Student Teaching, the number of candidates in each group differed slightly (Practicum $n = 28$; Student Teaching $n = 26$).

To address the second research question (whether K-12 students met their specific objectives), we again conducted separate analyses for Practicum ($n = 28$) and Student Teaching ($n = 26$). For this analysis, we compared the number of K-12 students per candidate who met their instructional objectives in the Learning Sequence with the number of students taught by each candidate. In addition, the total number of K-12 students who met their instructional objectives was divided by the total number of K-12 students across all candidates. This provided the overall percentage of K-12 students who met their objectives across all candidates. A dependent-sample t test was used to analyze each semester independently.

For the third research question (comparing the learning gains as candidates moved from Practicum to Student Teaching), 24 candidates comprised the cohort who participated in *both* Practicum and Student Teaching *during consecutive semesters*—that is, without breaks due to additional course requirements or other interruptions. With this cohort, we first analyzed the pretest–posttest learning gains in Practicum and Student Teaching separately, using a dependent-samples t test. Next, we analyzed the change from Practicum to Student Teaching using a dependent-samples t test. To analyze the impact on meeting their instructional objectives, we repeated this process, using a dependent-samples t test for the separate Practicum and Student Teaching students, and to analyze the change from Practicum to Student Teaching. For both analyses of the change from Practicum to Student Teaching, we confirmed the findings using a one-way ANOVA.

The fourth research question examined whether the type of lesson delivered by teacher candidates (i.e., academic vs. functional skills) might predict K-12 students' learning. For this question, only the Student Teaching experience was examined (n = 26). To analyze data for this question, an independent-samples t test was conducted to establish whether a significant difference was found between academic and functional lesson types for two types of K-12 student progress: (a) pre–post learning gains and (b) the percentage of K-12 students who met their specific learning objectives.

For the fifth research question (whether teacher candidates' GPAs might predict K-12 student learning gains), we again limited the analysis to candidates in Student Teaching. First, candidates were grouped by their cumulative GPAs for all of their special education coursework. Candidates' GPAs ranged between 3.25 and 4.0 on a 4-point scale. Next, candidates were ranked from the highest to lowest GPAs, and assigned to top, middle, and bottom GPA groups. The GPA for the top group ranged from 4.0 to 3.79 (n = 8); the GPA for the middle group ranged from 3.77 to 3.64 (n = 9); and the GPA for the bottom group ranged from 3.64 to 3.25 (n = 9). Finally, an independent-samples t test was conducted to examine whether there was a significant difference between the candidates' GPA groupings in relation to K-12 students' learning gains as measured by their pre–post change scores and the percentage of students who met their specific learning objectives.

Results

The results for the first research question are found in Table 1. K-12 students' learning gains in the teacher candidates' Practicum showed statistically significant ($p < .001$) pretest to posttest changes (M = 36.94, 95% confidence interval [CI] = [26.89, 46.99], SD = 23.79). These learning gains showed an effect size of d = 1.10 indicating a strong difference, where students' posttest scores increased by at least 1.1 standard deviations. During Student Teaching, K-12 students also demonstrated statistically significant ($p < .001$) pretest to posttest changes (M = 45.76, 95% CI = [34.99, 56.54], SD = 25.50), with an effect size of d = 1.80. This effect size also indicated a strong difference, where students' posttest scores increased by at least 1.8 standard deviations.

The results for the second research question are presented in Table 2. The number of K-12 students taught by teacher candidates during their Practicum ranged from 1 to 26 (M = 4.67, 95% CI = [2.77, 6.58], SD = 4.91); a statistically significant ($p < .001$) number of these students met their learning

Table 1. K-12 Student Learning Gains as Measured by Pretest and Posttest Scores During Practicum and Student Teaching.

Variable	*M* (%)	*SD*	*t*	*df*	Significance*	Cohen's *d*
Practicum[a]						
Pretest	47.54	23.75	6.73	27	.000	1.10
Posttest	80.13	19.91				
Change	32.59	25.64				
Student Teaching[b]						
Pretest	34.39	20.28	9.49	25	.000	1.80
Posttest	82.14	18.29				
Change	47.75	25.65				

Note. Cohen's (1988) measure of sample effect size for comparing two sample means, using a dependent-samples *t* test.
[a]Candidate $n = 28$; K-12 student $n = 131$.
[b]Candidate $n = 26$; K-12 student $n = 122$.
*$p < .05$ (two-tailed).

objectives ($M = 2.75$, 95% CI = [1.81, 3.69], $SD = 2.41$). During Student Teaching, the range of K-12 students participating in instructional activities per candidate was 3 to 8 ($M = 4.69$, 95% CI = [4.15, 5.25], $SD = 1.38$). A statistically significant ($p < .001$) number of these K-12 students met their learning objectives ($M = 3.23$, 95% CI = [2.59, 3.87], $SD = 1.58$). The effect size during Practicum ($d = 0.814$), and the effect size during Student Teaching ($d = 0.716$) indicate that the number of students meeting the learning objectives can be attributed to the K-12 students' participation in the teacher candidates' Learning Sequences. In aggregate, the K-12 students' learning gains resulted in their meeting specific objectives during both the candidates' Practicum and Student Teaching.

The results for the third research question are summarized in Table 3. The top section of Table 3 summarizes K-12 students' learning gains from pretest to posttest scores during Practicum ($M = 36.94$, 95% CI = [26.89, 46.99], $SD = 23.79$), and during Student Teaching ($M = 45.76$, 95% CI = [34.9, 56.54], $SD = 25.51$). These learning gains were statistically significant and showed strong effect sizes during Practicum ($p > .01$; $d = 1.79$) and Student Teaching ($p > .01$; $d = 2.34$). These findings are consistent with those reported previously in Table 1, with slight variations in candidate and student sample sizes. When examining the difference in learning gains as candidates progressed from Practicum to Student Teaching, results indicate an increase in student gains ($M = 8.83$, 95% CI = [5.45, 23.12], $SD = 33.83$), although these gains were not statistically significant, and the effect size was minimal ($p > .214$;

Table 2. K-12 Students' Learning Gains Resulting in Meeting Specific Objectives From the Learning Sequence.

Variable	*n*	*M*[a]	*SD*	*t*	*df*	Significance*	Cohen's *d*
Practicum[b]							
Students total	131	4.67	4.91	2.82	27	.007	0.814
Number (and %) who met objective	77 (58.7)	2.75	2.41				
Student Teaching[c]							
Students total	122	4.69	1.38	4.97	25	.000	0.716
Number (and %) who met objective	84 (68.9)	3.23	1.58				

Note. Cohen's (1988) measure of sample effect size for comparing two sample means.
[a]Mean number of K-12 students per teacher candidate.
[b]Candidate $n = 28$.
[c]Candidate $n = 26$.
*$p < .05$ (two-tailed).

$d = 0.358$). (Note: The results of a one-way ANOVA were consistent with those of the *t* test.)

The second variable addressed by Research Question 3, the percentage of K-12 students who met their specific learning objectives, is found in the bottom section of Table 3. The results showed statistically significant numbers of K-12 students met their learning objectives during Practicum ($M = 70.10$, 95% CI = [54.86, 83.95], $SD = 37.16$) and Student Teaching ($M = 64.85$, 95% CI = [51.79, 77.63], $SD = 32.70$), with strong effect sizes ($p > .01$, $d = 1.79$ during Practicum; and $p > .01$; $d = 2.34$ during Student Teaching). These results also were consistent with the findings reported in Table 2 based on variations in candidate and student sample sizes. When examining the findings as candidates progressed from Practicum to Student Teaching, results indicated a decrease in the percentage of K-12 students who met their learning objectives. However, this finding was not statistically significant ($M = 5.25$, 95% CI = [17.15, 27.65], $SD = 53.04$, $p > .632$) and showed a minimal effect size ($d = 0.149$). (Note: The results of a one-way ANOVA are consistent with those of the *t* test.)

The results for the fourth research question are found in Table 4. The top half of Table 4 summarizes K-12 students' learning gains from pretest to posttest score. Although K-12 students who participated in the academic lessons showed more improvement than those who participated in functional lessons, these differences were not statistically significant ($t = -1.173$, degrees of freedom [*df*] = 10, $p > .01$). Similarly, K-12 students who received the academic lessons also exceeded those who received the functional

Table 3. Candidates' Impact on K-12 Student Learning From Practicum to Student Teaching.

Variable[a]	*M*	*SD*	*t*	*df*	Significance*	Cohen's *d*
K-12 students' pre–post change scores						
Practicum percentage points gained	36.94	23.79	7.60	23	.000	1.786
Student Teaching percentage points gained	45.76	25.51	8.79	23	.000	2.339
Percentage points gained from Practicum to Student Teaching	8.83	33.83	1.28	23	.214	0.358
K-12 students who met learning objectives						
Met objective during Practicum[b]	70.10%	37.16	9.24	23	.000	0.516
Met objective during Student Teaching[c]	64.85%	32.70	9.715	23	.000	1.042
Change in meeting objective from Practicum to Student Teaching	−5.25%	53.04	0.485	23	.632	0.149

Note. Cohen's (1988) measure of sample effect size for comparing two sample means based upon the *t* statistic.
[a]Candidate sample size of $n = 24$ represents the cohort moving from Practicum to Student Teaching.
[b]K-12 student $n = 121$.
[c]K-12 student $n = 115$.
*$p < .05$ (two-tailed).

lessons (see the bottom half of Table 4), but these differences also were not statistically significant ($t = -0.911$, $df = 9$, $p > .01$).

Table 5 summarizes K-12 student learning gains during Student Teaching for teacher candidates grouped by their GPAs in the major (Research Question 5). For all candidates, the students' mean pretest–posttest learning gain was 47.77 percentage points (top section of Table 5). When students' pre–post gains were examined by candidates' GPA group assignments, candidates with GPAs in the top 1/3 showed a mean gain of 53.75 ($SD = 28.75$). Candidates in the middle and bottom GPA rankings showed gains of 45 ($SD = 14.92$) and 45.22 ($SD = 32.55$) percentage points, respectively. This resulted in a

Table 4. Impact of Lesson Type on K-12 Student Learning During Student Teaching.

Type of lesson delivered by teacher candidate[a]	*M*	*SD*	*t*	*df*	Significance*	Cohen's *d*[b]
K-12 students' pre–post change scores (percentage points)						
Academic[c]	51.42	25.05				
Functional[d]	37.86	26.54				
Difference	13.56	NA	−1.173	10	.267	0.525
Percentage of K-12 students who met learning objectives						
Academic[c]	71.79	30.68				
Functional[d]	56.86	38.30				
Difference	14.67	NA	−0.911	9	.386	0.430

[a]Teacher candidate $n = 26$.
[b]Cohen's (1988) measure of sample effect size for comparing two sample means based upon the t statistic.
[c]K-12 student $n = 86$.
[d]K-12 student $n = 36$.
*$p < .05$ (two-tailed).

statistically significant difference between the top and middle GPA groups ($t = 2.42$, $df = 11$, $p < .05$). Results *did not* reveal statistically significant differences between the top and bottom GPA groups ($t = 1.76$, $df = 15$, $p > .05$), or the middle and bottom groups ($t = −0.06$, $df = 11$, $p > .05$).

Finally, the bottom section of Table 5 shows K-12 learning gains as measured by meeting their learning objectives. Across all candidates, 80 of the 122 students (65.57%) met their specific learning objectives. For candidates in the top GPA group, 76.13% (*SD* = 30.57) of their K-12 students met their learning objectives. For candidates in the middle and bottom GPA groups, 58.55% (*SD* = 25.93) and 69% (*SD* = 41.19) of their K-12 students met their learning objectives. This resulted in a statistically significant difference between the top and middle GPA groups ($t = 3.9$, $df = 14$, $p < .01$). Results *did not* reveal statistically significant differences between the top and bottom GPA groups ($t = 1.25$, $df = 14$, $p > .05$), or the middle and bottom groups ($t = −1.93$, $df = 14$, $p > .05$).

Discussion

The purpose of this study was to explore whether instruction delivered by candidates in a TPP would result in measurable learning gains in their K-12

Table 5. Summary of Differences in Teacher Candidates by GPA Grouping During Student Teaching.

		Difference scores		
GPA for teacher candidates	K-12 student change	Top to middle GPA group comparison	Top to bottom GPA group comparison	Middle to bottom GPA group comparison
K-12 students' pre–post change scores (percentage points)				
All candidates[a]	47.77			
Top 1/3[b]	53.75	—	—	
Middle 1/3[c]	45.0	—		—
Bottom 1/3[d]	45.22		—	—
Difference		8.75	8.53	0.22
		$t(11) = 2.42^{*}$	$t(15) = 1.76$	$t(11) = -0.06$
Percentage of K-12 students who met learning objectives				
All candidates[a]	65.57			
Top 1/3[b]	76.13	—	—	
Middle 1/3[c]	58.55	—		—
Bottom 1/3[d]	69.0		—	—
Difference		17.58	7.13	10.45
		$t(14) = 3.9^{**}$	$t(14) = 1.25$	$t(14) = -1.93$

Note. GPA = grade point average.
[a]Candidate $n = 26$, K-12 student $n = 122$.
[b]K-12 student $n = 33$.
[c]K-12 student $n = 45$.
[d]K-12 student $n = 44$.
$^{*}p < .05$. $^{**}p < .01$.

students with disabilities. To assess learning gains, we examined student performance on learning objectives from the instructional Learning Sequences delivered *only by the candidates*. Results indicate that K-12 students demonstrated notable learning gains on two separate but related measures: (a) pretest to posttest gain scores and (b) the percentage of students who met the criteria established for mastery of their learning objectives. These results replicate the previous findings by Brady et al. (2016), who reported similar learning gains with three separate cohorts of graduate and undergraduate teacher candidates. Similar to the findings reported by Brady et al. (2016), the learning gains in *this study* were evident during the candidates' part-time Practicum experiences, as well as during their full-time Student Teaching experiences. These findings, however, extend the previous research by reporting findings that were not evident in the earlier work; when the candidates progressed from Practicum to their Student Teaching experiences, there was an increase in K-12 students' learning gains (i.e., their pre–post measures), but not in the percentage of students who met their specific learning objectives, and none of these changes were significant. Together, these two studies illustrate the ability of TPPs to demonstrate the capacity of teacher candidates to influence learning gains in K-12 students as a direct result of their instructional experiences with their students. In addition, the current study extends this line of research by showing the differential impact of candidates' part-time (Practicum) versus full-time (Student Teaching) instructional activities, and by showing that K-12 student learning may also be influenced by the context of the curriculum delivered to the students (i.e., academic content vs. instructional on functional skills).

A significant flaw exists in the logic that student performance on high-stakes tests can be attributed primarily to their teachers, and this has been pointed out by numerous researchers (AERA, 2015; Amrein-Beardsley, 2008; Ballou & Springer, 2015; Holdheide et al., 2010; Papay, 2011). Isolating the effects of an individual teacher on students who receive instruction from a multitude of different teachers is a serious challenge, particularly when these students are not assessed regularly on the actual content taught by that teacher (Steinbrecher et al., 2014). This is a frequent challenge when tracking learning in English Language Learners, students with disabilities, and non-fluent readers (Jones et al., 2013). In the present study, K-12 student performance was assessed *only on content that was actually delivered by the candidates*, and student performance was linked *only to the candidates who actually delivered that content*. As such, this study showed a *direct* link between student performance and candidates' teaching behavior. This direct link overcomes many of the issues that threaten the value-added logic applied to students whose standardized assessment performance does not accurately portray their learning, and whose learning characteristics require accommodations (e.g., non-standardized curriculum, learning support teachers, testing accommodations, and more frequent performance assessments).

One might expect that candidates would have a greater effect on students' learning during their final, full-time Student Teaching experiences, and some results did support that. The intensity of the instruction (90 student contact hours) delivered during Practicum is far less than during Student Teaching (600 student contact hours); intuitively one would expect a differential improvement in candidates' ability to affect students during Student Teaching. Second, the amount and type of feedback on candidates' performance changed as the candidates acquired more teaching experience. Practicum served as the candidates' first intensive field experience. Candidates received a minimum of seven formal observations from university supervisors and classroom teachers; after these observations, candidates received thorough feedback and mentoring from their classroom teachers and university supervisors, including coaching and specific prescriptive feedback immediately following the lesson that was observed. By the time candidates engaged K-12 students during Student Teaching, they had additional teaching practice, and typically developed more stable teaching patterns. Also, they participated in this experience on a full-time basis for 16 continuous weeks. As such, the candidates gained considerably more practice and feedback on their own teaching performance, and it is reasonable to expect that they would be more effective teachers, with a greater impact on student learning. That is, as candidates gained both experience and coaching across two semesters, one would expect that student learning gains would increase substantially. The finding that the percentage of students who met their objectives *decreased* from Practicum to Student Teaching was unexpected when considering the factors above, and requires further attention in subsequent investigations. Perhaps this finding is related to differences in the nature of the students or the types of classes in which the candidates taught. A post hoc review of K-12 student characteristics shows two differences in the student populations during Practicum and Student Teaching. Students during the Practicum experience were identified primarily as having emotional or behavioral disorders, or learning disabilities, and were taught in larger classroom settings. By contrast, students during the Student Teaching experience had a higher incidence of autism spectrum disorders and intellectual disabilities, and were typically taught in smaller classes. In addition, K-12 students in Practicum were somewhat older (M age = 13.1; SD = 3.93) than those during Student Teaching (M age = 10.7; SD = 4.14). Thus, the student population during Practicum could be described as having milder academic disabilities and were older, where the students in Student Teaching would be classified as having more moderate to severe disabilities and were younger.

The difference between the contextual nature of the lessons indicated there was no statistically significant difference between academic and functional instruction. This could indicate and be a result of the fact that each Learning Sequence specifically addressed one objective. Whether or not the Learning Sequence was targeting an academic or functional skill, each was designed to be accomplished over a minimum of eight lessons. Teacher candidates had to ensure that each Learning Sequence objective could be achieved in that time period. However, there were apparent differences between the mean pre–post change scores between academic and functional skill acquisition. In addition, there is an apparent difference in the percentage of K-12 students who met their learning objectives in academic and functional skill instruction. Although it is not statistically significant, there is an obvious difference between the two categories. It could be that most students who received instruction in functional skills have more significant disabilities than those who received academic skills. Due to that reason, significant learning gains may take longer to develop than what was demonstrated across the Learning Sequence. Also, the smaller sample size (teacher candidate n = 26) could have contributed to the lack of statistical significance between the two groups.

Finally, there were varying results regarding teacher candidates' performance and their GPAs in the major, an area of inquiry not explored in previous research by Brady et al. (2016). Although there were some differences among the groups, there was no consistency in these findings. To understand these findings, it is important to keep in mind that the pool of candidates had to maintain high GPAs to matriculate into Student Teaching. It could be that only minimal differences exist among candidates because the grade differences among them are minute; all candidates were within 0.75 points on a 4.0 scale. In future studies, it might be more valuable to examine grades and performance in particular courses (e.g., methods, assessment, or classroom management courses) rather than the overall GPA in the major field of study.

Limitations and Future Research

Although these findings were encouraging and suggest an alternate model of linking K-12 student learning to candidates' performance in a TPP, this study has at least four limitations that must be considered. First, the participants in this study represented all teacher candidates in a single TPP, during a single, two-semester academic year. Because we did not study multiple cohorts across different TPPs, the generalizability of the findings is obviously limited. Furthermore, these candidates were all undergraduate students matriculating toward the bachelor's degree in special education. No attempt was made to include graduate students or those enrolled in general education programs. Second, the type of instructional delivery provided by the candidates was drawn from an explicit, direct instruction model characteristic of special education teachers. This model of instruction lends itself to an orderly scope and sequence of curricular objectives, and includes frequent curriculum-based assessment. Obviously, this instructional model does not represent the methods of instruction provided in some TPPs. Instructional models that rely more on discovery or project-based learning could find the measures of learning gains in this study inconsistent with their own measures and methods of instruction.

The third limitation is based on the small sample size of this cohort, and the state certification standards that demand that candidates gain teaching experience across a broad range of ages and types of disabilities. These two pragmatic circumstances limited our ability to probe specific learning outcomes for any given age of students, grade level, disability characteristics, or curriculum and content areas. It is possible that candidates showed more or less effect on specific ages or grade levels of students, or in the many different content areas that they taught (e.g., early reading, mathematics word problems, personal hygiene and care skills, social problem-solving routines). The limited number of candidates and their distribution across the student and curricular variables simply precluded any meaningful analysis of differential effects.

Finally, the design of the preparation program requires that candidates' schools, settings, K-12 students, and classroom teachers differ from their Practicum to Student Teaching. Although this assures that candidates gain diverse teaching experiences, it also results in limited time to assess candidates' impact on the students whom they teach. Ideally, any assessment of candidates' impact on K-12 student learning would be based on longer periods of time; this would reduce the probability that candidate impact is a novelty effect and allow investigators to examine a longer term impact of their teaching.

These limitations and other questions that arose from our results provide several possibilities for future investigations. First, there is an obvious need to replicate this study if candidate delivery of instruction is to become a viable alternative to the VAM approaches based on standardized high-stakes student assessments. Research with elementary education teacher candidates, in co-teaching arrangements, and with students who are just learning to use English for instruction are obviously high priority for replications. Among special education teachers in preparation, it would be valuable to distinguish between teacher candidates who deliver functional curriculum to students with moderate to severe disabilities versus those who primarily deliver academic curriculum in either resource settings or inclusive classrooms.

A second area for future research involves the metrics collected and reported by the candidates. The data collected in this study are drawn from information required by the TPP, reflecting faculty and school district input on effective teaching practices. In Practicum, candidates were asked to report change scores based on group averages as the student performance measure instead of individual student scores. Obviously, this limited the analyses used in the study. Given the wide distribution of student performance on their assignments, and the resultant multiple outliers in the data, a better analytic tool might include analysis of medians as measures of central tendency. The TPP revised the candidate reporting requirement mid-year, and subsequently, candidates began to report individual pretest and posttest scores. This change will allow investigators to explore other approaches to the data analysis in the future, and will expand the data analyses to better derive the nuanced findings.

A third future research focus involves an expansion of variables that might *predict* whether candidates are likely to have a positive impact on their students' learning. Serious research on effective teaching spans at least four decades, and numerous variables have been identified that directly or indirectly promote student learning (Berliner, 1976; Goe, Bell, & Little, 2008). These variables generally fall into one of three clusters: (a) student conditions, (b) teacher variables, and (c) environmental conditions (Creemers & Reezigt, 1996). In the current study, we selected only a handful of predictor variables, and the results showed that this is a valuable area to pursue further. We are currently preparing to follow up this line of inquiry by examining specific teacher performance characteristics measured by a teacher observation scale. Such items reflect particular teacher planning and delivery behaviors, and could be much more predictive than broader indices such as GPA or lesson types.

Conclusion

TPPs continue to struggle with regulatory and accreditation demands that base their program evaluations on K-12 learning gains of their candidates. Like many legislated efforts to promote accountability in public education, applying VAM logic to teacher evaluations and TPPs has resulted in numerous unintended negative consequences (Brady et al., 2014; Moore Johnson, 2015; Morganstein & Wasserstein, 2014). These negative consequences include new obstacles to placing candidates in authentic field experiences, as classroom teachers become less likely to accept teacher candidates due to concerns that if K-12 student performance does not show rapid gains, their own evaluations (and promotions and salaries) could be affected negatively. Another unintended negative outcome of the original VAM efforts is a decreased willingness to teach students who require additional effort, accommodations, and support (Berliner, 2013). For TPPs, this narrows the pool of school placements even further, as classroom teachers decide that neither teacher candidates nor K-12 students with more complex learning conditions are welcomed or included in their classes. A third set of unintended consequences is a direct outcome of the complexity of most current VAMs. As teachers and TPPs wrestle with increasingly complex VAM challenges, the financial and human costs of implementing these systems diverts significant resources from instruction (Darling-Hammond, 2015; McCullough et al., 2015; Morganstein & Wasserstein, 2014).

These and other policy and practice issues will continue to vex teacher educators as they struggle to create and apply a rational VAM for TPPs. We believe that the VAM investigated in this study provides a rational alternative to the current models based on high-stakes standardized assessments that face basic validity and reliability challenges, and responds to recent calls for alternative VAMs (AERA, 2015).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

American Association of Colleges for Teacher Education. (n.d.). *About edTPA*. Retrieved from http://edtpa.aacte.org/about-edtpa

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*, 448-452.

Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher*, *37*, 65-75.

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, *44*, 77-86.

Berliner, D. C. (1976). Impediments to the study of teacher effectiveness. *Journal of Teacher Education*, *27*, 5-13.

Berliner, D. C. (2013). Problems with value-added evaluations of teachers? Let me count the ways! *The Teacher Educator*, *48*, 235-243. doi:10.1080/08878730.2013.827496

Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Press*, *116*(1) Article 17293. Retrieved from http://www.tcrecord.org/library

Brady, M. P., Duffy, M. L., Hazelkorn, M., & Bucholz, J. (2014). Policy and systems change: Planning for unintended consequences. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *87*, 102-109.

Brady, M. P., Heiser, L. A., McCormick, J. K., & Forgan, J. (2016). Adding value-added models to teacher preparation programs: Validity and reliability threats, and a manageable alternative. *The Educational Forum*, *80*, 339-352.

Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability: Measurement challenges for students with disabilities and recommendations for research. *Educational Researcher*, *39*, 537-544.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Creemers, B. P. M., & Reezigt, G. J. (1996). School-level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*, *7*, 197-228.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44*, 132-137.

Elliott, S. N., Roach, A. T., & Kurz, A. (2014). Evaluating and advancing the effective teaching of special educators with a dynamic instructional practices portfolio. *Assessment for Effective Intervention*, *39*, 83-98.

Floden, R. (2012). Teacher value added as a measure of program quality: Interpret with caution. *Journal of Teacher Education*, *63*, 356-360.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room Value Added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*, 96-104.

Hamilton, R., & Brady, M. P. (1991). Individual and classwide patterns of teachers' questioning in mainstreamed social studies and science classes. *Teaching and Teacher Education*, *7*, 253-262.

Hill, H. C., Kapitula, L., & Umland, K. A. (2011). Validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*, 794-831.

Holdheide, L., Goe, L., Croft, A., & Reschly, D. (2010). *Challenges in evaluating special education teachers and English language learner specialists*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/July2010Brief.pdf

Jones, N. D., & Brownell, M. T. (2014). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention*, *39*, 112-124.

Jones, N. D., Buzick, H. M., & Turkan, S. (2013). Including students with disabilities and English learners in measures of educator effectiveness. *Educational Researcher*, *42*, 234-241.

Lincove, J. A., Osborne, C., Dillon, A., & Mills, N. (2014). The politics and statistics of value-added modeling for accountability of teacher preparation programs. *Journal of Teacher Education*, *65*, 24-38.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stetcher, B. M., Le, V., & Martinez, F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, *44*, 47-67.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, *4*, 572-606.

McCullough, M., English, B., Angus, M. H., & Gill, B. (2015). *Alternative student growth measures for teacher evaluation: Implementation experiences of early-adopting districts* (REL 2015–093). Washington, DC: Regional Educational Laboratory Mid-Atlantic, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/edlabs

Moore Johnson, S. (2015). Will VAMs reinforce the walls of the egg-crate school? *Educational Researcher*, *44*, 117-126.

Morganstein, D., & Wasserstein, R. (2014). ASA statement on value-added models. *Statistics and Public Policy*, *1*, 108-110. doi:10.1080/2330443X.2014.956906

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Education Research Journal*, *48*, 163-193.

Pecheone, R. L., Shear, B., Whittaker, A., & Darling-Hammond, L. (2013). *2013 edTPA field test: Summary report*. Stanford, CA: Stanford Center for Assessment, Learning and Equity.

Steinbrecher, T. D., Selig, J. P., Cosbey, J., & Thorstensen, B. I. (2014). Evaluating special educator effectiveness: Addressing issues inherent to value-added modeling. *Exceptional Children*, *80*, 323-336.

Swank, P. R., Taylor, R. D., Brady, M. P., & Freiberg, J. (1989). Sensitivity of classroom observation systems: Measuring teacher effectiveness. *Journal of Experimental Education*, *57*, 171-186.

Author Biographies

Michael P. Brady is a Professor and the Chair of the Department of Exceptional Student Education at Florida Atlantic University. His research interests include policies that affect teacher preparation, and behavioral teaching strategies for people with developmental disabilities.

Katie M. Miller is an Assistant Professor of Exceptional Student Education at Florida Atlantic University. Her research interests include writing and learning strategies, and teacher preparation.

Jazarae McCormick is a Doctoral Scholar in Exceptional Student Education at Florida Atlantic University. In addition to her research interests in teacher preparation, she studies federal policies that promote leadership training in education, and strategies that empower adults with developmental disabilities to participate meaningfully in their own educational opportunities.

Lawrence A. Heiser is a Senior Instructor in Exceptional Student Education at Florida Atlantic University. In addition to coordinating field experiences for teacher education majors, he is active in promoting professional development and mentoring in teachers.