

USING EMPIRICAL EVIDENCE TO EVALUATE TEACHER PREPARATION PROGRAMS: A CASE STUDY FOR CAEP ACCREDITATION

Tingting Xu, Ph.D.

Stephen F. Austin State University

Tracey Covington Hasbun, Ph.D.

Stephen F. Austin State University

Abstract

Guided by Standards 4 and 5 of the Council for the Accreditation of Educator Preparation (CAEP, 2020), the intent of the study was to provide a comprehensive evaluation of the quality of one teacher preparation program in rural East Texas. A case study with a mixed-method design was used to examine empirical data from a broad set of measures. Based on feedback from mentor cooperating teachers, recent graduates, and school principals, strengths and weaknesses of the teacher preparation program were identified. Results showed that our teacher candidates were well prepared with a good understanding of individual differences and the diverse backgrounds of learners. The candidates were well prepared to create environments that support individual and collaborative learning, encourage positive social interaction, and facilitate active engagement in learning. Results also revealed a need to better prepare teacher candidates with skills in assessment and leadership, and to consider ways to strengthen proficiency in classroom management.

Keywords: Teacher preparation, CAEP accreditation, case study

There is growing attention on the effectiveness of teacher preparation programs (TPPs) in preparing quality teachers. Teacher preparation programs are frequently asked to determine how effective their graduates are based on several factors such as job placement, retention rates, student learning, and graduates' perceptions on performance and effectiveness of their TPPs (Henry et al., 2012; Monk 2015; U.S. Department of Education, 2011, as cited in Coggshall et al., 2012). Investigating this topic is important because "Too many beginning teachers report that they do not feel well-prepared when they enter the classroom, and their supervisors often agree" (Levine, 2006 as cited in Coggshall et al., 2012, p. 3). Only as few as 20% of first-year teachers indicated they felt they were well-prepared to choose and modify curriculum, handle classroom management, and assess the students under their care (National Center for Education Statistics, 2011 as cited in Coggshall et al., 2012). Stakeholders have called for

increased research to improve the quality and effectiveness of teacher education programs (Bartell et al., 2018).

To evaluate and make decisions on program effectiveness, emphasis has been placed on collecting empirical evidence for improvement in the TPPs (Crowe, 2010; Ludlow et al., 2011). This is guided by the Council for the Accreditation of Educator Preparation (CAEP) Standard 5: Provider quality, continuous improvement, and capacity (CAEP, 2020), which suggests that a quality assurance system relies on a variety of comprehensive measures to ensure the continuous improvement of the TPPs (Ruben, 2010). For example, areas to measure program quality for improvement could be: a) candidates' characteristics, b) their knowledge, skills, and professional dispositions, and c) their impact on students' learning, which was recognized as value-added assessment (Bransford et al., 2005; Zeichner & Conklin, 2005). However, value-added assessment alone does not provide

detail as to what to improve (Noell et al., 2019) given the limitations of the grade or subject area tested (American Education Research Association [AERA], 2015; Henry et al., 2012). Classroom observation evaluation ratings of teacher performance are more informative when making decisions of hiring a new teacher, as indicated by school administrators, (Goldring et al., 2015), and can be a valuable component of TPP evaluation (Bastian et al., 2018). Given the complexity in the process of evaluating the effectiveness of TPPs, “attention should be given to incorporating a broader set of teacher performance measures” (Henry et al., 2012 p. 351).

Based on the suggestions from the current literature, as well as CAEP standard 4: 1) 4.2-Indicators of Teacher Effectiveness, 2) 4.3-Satisfaction of Employers, and 3) 4.4-Satisfaction of Completer (CAEP, 2020), comprehensive empirical data were collected from a broad set of outcomes with the intention to provide a systematic evaluation of the quality of the TPP. In the current study, investigators focused on beginning teachers with less than three years of teaching experience, given the fact that the influence of teacher education preparation might be weakened over time. Examination of similarities and differences between candidate performance during clinical teaching and teaching performance at the conclusion of at least one year of serving as a teacher of record in a classroom setting was conducted through multiple resources. These sources served to answer the following research questions:

1. How well did we prepare our teacher candidates?
2. What were the similarities and differences in their teaching before and after graduation?
3. What can be improved in our teacher education program?

Methodology

Participants

Graduates who successfully completed the Early Childhood through 6th Grade (EC-6) and Middle Level Grade (4-8) programs offered through the TPP and who had also served between one to three years as a teacher of record in a classroom setting served as participants. To begin with, a convenience sample of participants was identified by faculty members and program coordinators from within the Department. Faculty members nominated former teacher candidates with whom they had worked and

who also agreed to participate. The final group of participants included ten graduates from the EC-6 and 4-8 programs, two males and eight females, who were African American, Hispanic, or White. These beginning teachers were placed in Pre-K, K, 1st, 3rd, 5th, 6th, and 8th grade classrooms. In addition, seven principals participated in this study. The principals were predominantly female, most of which were African American and White, and all served as administrators on the teachers’ home campuses. Faculty members and mentor cooperating teachers (i.e., classroom teachers who supervised teacher candidates in their clinical teaching experience) also completed evaluations which provided data for this study. However, they were not considered primary participants.

Data Collection and Analysis

Data collection took place in two different periods. The first round of data was collected during the clinical teaching period. The second round of data was collected near the completion of at least one year of serving as a teacher of record in a classroom setting. Data were collected from multiple instruments as described in the following.

The Program Evaluation Survey

This survey included ten statements that were directly modified from the Interstate Teacher Assessment and Support Consortium (*InTASC*) standards by the Council of Chief State School Officers (2013). This survey was completed by the participants as they finished the TPP program just prior to graduation, and then, again, at the conclusion of at least one year of serving as a teacher of record in a classroom setting. The survey was also completed by the school principal or administrator responsible for the supervision of the participant, near the end of at least one year of service as the teacher of record.

The Candidate Performance Evaluation Survey

This instrument included ten areas of the *InTASC* standards by the Council of Chief State School Officers (2013). This instrument was completed by the mentor cooperating teachers upon candidates’ completion of the program, just prior to graduation. The same survey was also completed by the graduates at the conclusion of at least one year of serving as a teacher of record in a classroom setting.

Formal Observations Using the Texas Teacher Evaluation and Support System (T-TESS) Evaluations

The Texas Teacher Evaluation and Support System (T-TESS) Evaluations (Texas Education Agency, 2016) is a state-approved educator evaluation measure. It was completed by district administrators and university faculty through face-to-face or virtual observation of the participant delivering instruction in the classroom setting using the T-TESS. It is important to note that, at one campus, a T-TESS was not used, but a state-approved measure was used to conduct a formal evaluation of the teacher of record.

Interviews with Beginning Teachers and Principals

Semi-structured interview questions were used, each of which was associated with one of the T-TESS domains. The interviews took place at the conclusion of at least one year of serving as a teacher of record in a classroom setting and were led by university faculty.

Student Performance Data

These data were mostly from the State of Texas Assessments of Academic Readiness (*STAAR*[®]) tests or the STAAR practice tests (TEA, 2020), which ranged from 3rd grade to 8th grade and subjects included math, reading, and science. First-grade students' performance was evaluated using the Neuhaus Reading Readiness, the Fountas & Pinnell Benchmark Assessment, and Renaissance Reading and Math STAAR tests. Pre-K and Kindergarten student performance data were limited and not included.

Quantitative data were analyzed using SPSS 25.0 (IBM Corp., Armonk, NY). For all quantitative data collected from instruments, descriptive statistics were used to present the average in the responses. Non-parametric Wilcoxon Signed Rank tests were used to compare the similarities and differences between candidates', teachers', and principals' responses. For all interview data, open coding (Strauss, 1987) was used to seek patterns and themes. Data were transcribed, and then the two researchers coded all data separately. Interrater reliability was analyzed during the coding process, and 80.5% consensus was reached for the teachers' interview data, and 82.8% consensus was reached on principals' interview data. Researchers then came to agreement on each individual theme.

Results

Results are presented in a format to clearly answer the research questions. First, we wanted to know how our program prepared the teacher candidates? The results from the Program Evaluation Survey by candidates, teachers, and principals showed that they all believed that our program prepared quality teacher candidates. In detail, candidates believed they were fully prepared, shown through average scores of 3.0. When becoming teachers in the classroom, they believed they were prepared with averages ranging from 2.4 to 2.8. Principals rated similarly with averages from 2.4 to 2.85 in these statements. Non-parametric Wilcoxon Signed Rank tests showed no significant difference in program evaluations rated by candidates, teachers, and principals, indicating that they were aligned in their opinions about the quality of the candidates our program prepared.

When examining each statement, candidates believed they were fully prepared to teach as they rated 3.0 out of 3.0 in every statement. When they became a teacher, they believed they were fully prepared in response to the following statement: 1) To demonstrate an understanding of individual differences, diverse cultures, and communities to ensure inclusive learning environments that enable each learner to meet high standards. They felt a lack of preparation in these two lowest-rated statements: 2) understanding of how to connect concepts and use differing perspectives to engage learners in critical thinking, creativity, and collaborative problem solving related to authentic local and global issues, and 3) understanding and use of multiple methods of assessment to engage learners in their own growth, to monitor learner progress, and guide the teacher's and learner's decision making.

Principals believed these teachers were fully prepared in statement 1): To work with others to create environments that support individual learning, collaborative learning, encourage positive social interaction, active engagement in learning, and self-motivation. However, at the same time, teachers needed more preparations in statement 2): Seeking appropriate leadership roles and opportunities to take responsibility for student learning; to collaborate with learners, families, colleagues, other school professionals, and community members to ensure learner growth; and to advance the profession. Details are presented in Table 1.

Table 1
Program Evaluation by Beginning Teachers & Principals

Program Evaluation	n	By Beginning Teachers		By Principals	
		Mean	Std. Dev.	Mean	Std. Dev.
How well did the candidate demonstrate understanding of how learners grow and develop, recognizing that patterns of learning and development vary individually within and across the cognitive, linguistic, social, emotional, and physical area, and designs and implements developmentally appropriate and challenging learning experiences? INTASC-2010.1	10	2.800	.422	2.500	.527
How well did the candidate demonstrate understanding of individual differences, diverse cultures, and communities to ensure inclusive learning environments that enable each learner to meet high standards? INTASC-2010.2	10	2.900	.316	2.600	.516
How well did the candidate work with others to create environments that support individual learning, collaborative learning, encourage positive social interaction, active engagement in learning, and self-motivation? INTASC-2010.3	10	2.800	.422	2.850	.337
How well did the candidate demonstrate understanding of the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and creates learning experiences that make these aspects of the discipline accessible and meaningful for learners to assure mastery of the content? INTASC-2010.4	10	2.700	.483	2.700	.483
How well did the candidate demonstrate understanding of how to connect concepts and use differing perspectives to engage learners in critical thinking, creativity, and collaborative problem solving related to authentic local and global issues? INTASC-2010.5	10	2.400	.699	2.700	.483
How well did the candidate demonstrate understanding and use of multiple methods of assessment to engage learners in their own growth, monitor learner progress, and guide the teacher's and learner's decision making? INTASC-2010.6	10	2.400	.699	2.500	.527
How well did the candidate plan instruction that supports every student in meeting rigorous learning goals by drawing upon knowledge of content areas, curriculum, cross-disciplinary skills, and pedagogy, as well as knowledge of learners and the community context? INTASC-2010.7	10	2.600	.699	2.600	.516
How well did the candidate demonstrate understanding and use of a variety of instructional strategies to encourage learners to develop deep understanding of content areas, their connections, and build skills to apply knowledge in meaningful ways? INTASC-2010.8	10	2.700	.675	2.800	.422
How well did the candidate engage in ongoing professional learning, adapt practice to meet the needs of each learner, and to use evidence to continually evaluate my practice; particularly, the effect of my choices/actions on others (learners, families, other professionals, and the community)? INTASC-2010.9	10	2.500	.972	2.800	.422
How well did the candidate seek appropriate leadership roles and opportunities to take responsibility for student learning; to collaborate with learners, families, colleagues, other school professionals, and community members to ensure learner growth; and to advance the profession? INTASC-2010.10	10	2.500	1.080	2.400	.699

The students' performance data also demonstrated how our program prepared the teacher candidates. It was evident with the positive impact that our graduates had on students' academic achievement. Multiple sets of the same tests from the same teacher results showed increased passing scores on their academic performance tests. For example, 70% of the first-grade students reached the grade level of reading accuracy on the first test, and this accuracy increased to 95% in the second test. Of these students, 60% reached a score of 100 (the maximum score) in comprehension on the first test, but 90% of them reached the maximum score of

100 in comprehension on their second test. The same pattern was found for their Renaissance reading and math results, with test scores increasing in each test (see Figures 1 and 2). Students' growth was evident in 8th grade English language arts performance as scores went from 95% to 100% in passing rates for one group of students and 90% to 94% for another 8th grade class. Math test results and science test results were a bit lower, with the average passing rate of 66% for 3rd grade math tests and 69.97% for 6th grade science tests.

Figure 1
First Grade Renaissance Math Test Results

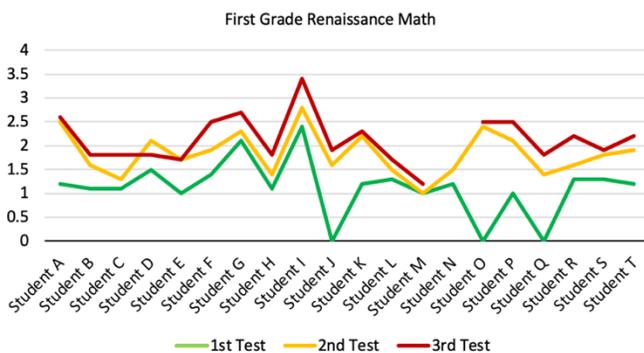
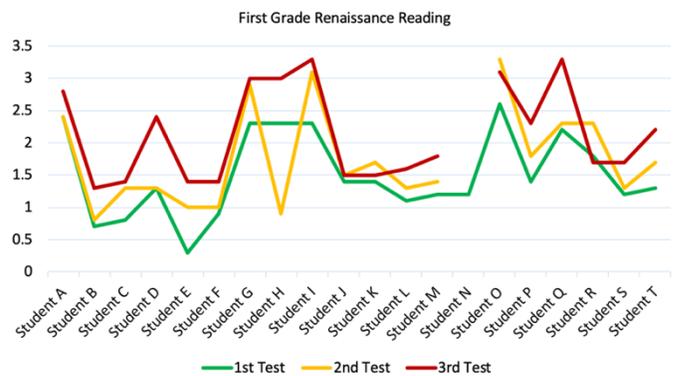


Figure 2
First Grade Renaissance Reading Test Results



The Similarities and Differences in Teaching Performance Before and After Graduation

For the Candidate Performance Evaluation Survey completed by the mentor cooperating teacher, the average scores ranged from 2.500 to 2.929, with SD ranging from .189 to .608 in these ten statements. Candidates received the highest ratings on these areas:

1. Understanding of how learners grow and develop, recognizing that patterns of learning and development vary individually within and across the cognitive, linguistic, social, emotional, and physical areas, and designs and implements developmentally appropriate and challenging learning experiences.
2. Understanding of individual differences, diverse cultures, and communities to ensure inclusive learning environments that enable each learner to meet high standards.

3. Working with others to create environments that support individual learning, collaborative learning, encourage positive social interaction, active engagement in learning, and self-motivation.

Our principal interview responses also echoed statements regarding strength in creating inclusive environments that support individual students and working with others to support learning. In particular, a principal stated that one of her new teacher's strengths was "her willingness and drive to become a great teacher. She has asked to meet each week with the curriculum AP before her team plans so that she can contribute to planning. She works with her team and the Curriculum Assistant Principal to plan lessons weekly."

Of a second teacher on her campus, the principal stated that she, too, "participates in PLC discussions each week to answer the questions: (1) What are you going to do for the

students who did not get it? (2) What are you doing for those who already know it?"

On another campus, a principal spoke to the positive environment that his novice teacher cultivated in his classroom and how this familial environment motivated students to engage in learning by saying:

"One of the things that make his room so fun almost is right at the beginning of the year he spends a lot of time creating their family and you know, and that's when you walk in at 8 in the morning, they're on the carpet together sharing about what they ate for dinner last night or you know, I mean anything. Somebody gave their dog a bath, you know, but they all know it, and they all are part of each other's lives, and that's their family during the day, and so he takes a lot of time to create that environment which goes right along with he's got the kids engaged in the problem, you know, and they're going to do whatever Mr. Smith (pseudonym) says because that's the family and that's the environment that he's created."

They received the lowest rating in the statement: Understanding and using multiple methods of assessment to engage learners in their own growth, monitor learner progress, and guide the teacher's and learner's decision making.

For the Teacher Performance Evaluation rated by principals, the average scores ranged from 2.286 to 2.857, with SD ranging from .378 to .756 for these ten statements. Graduates received the highest rating on: Engaging in ongoing professional learning, adapting practice to meet the needs of each learner, and using evidence to continually evaluate his or her practice; particularly, the effect of his or her choices/actions on others (learners, families, other professionals, and the community).

They received the lowest ratings on these areas:

1. Understanding of how learners grow and develop, recognizing that patterns of learning and development vary individually within and across the cognitive, linguistic, social, emotional, and physical areas, and designs and implements developmentally appropriate and challenging learning experiences.
2. Understanding and use of multiple methods of assessment to engage learners in their own growth,

monitor learner progress, and guide the teacher's and learner's decision making.

3. Seeking appropriate leadership roles and opportunities to take responsibility for student learning; to collaborate with learners, families, colleagues, other school professionals, and community members to ensure learner growth; and to advance the profession. This is also echoed in the Candidate Performance Evaluation Survey completed by mentor cooperating teachers, as well as the principal interviews, that some beginning teachers need support on assessing students and using students' data for planning.

At one elementary school, the principal stated that her newly hired teacher's "biggest weakness is assessment. That is her biggest concern. She wants to know how they are created and aligned to STAAR. She's uncertain how to create her own assessments that are aligned to the skills" that are to be learned.

Another principal noted a similar area of need in one of her teachers. Although she considered the new teacher to be proficient and, in informal conversations with the researcher, mentioned that she is really strong as a new teacher, in the formal interview, the principal stated that "the area that she needs to do some work on would be gathering and using formative assessments throughout the lesson cycle."

Correlation analysis showed the relationship between mentor cooperating teachers' rating on candidate performance and principals' rating on teacher performance were all positive, ranging from .167 to .750, with mostly median to high correlation. There was a significant positive correlation between mentor rating and principal rating on the statement of working with others to create environments that support individual learning, collaborative learning, encourage positive social interaction, active engagement in learning, and self-motivation? with $r = .881$, $p = .009$.

We compared the means of candidate performance evaluation by mentor teacher and teacher performance evaluation by principal using the non-parametric test. Results showed that there was a significant difference in statement of demonstrating understanding of how learners grow and develop, recognizing that patterns of learning and development vary individually within and across the

cognitive, linguistic, social, emotional, and physical areas, and designs and implements developmentally appropriate and challenging learning experiences? Principal rated this statement significantly lower than mentor teachers, with $Z = -2.121, p = .034$.

University supervisors utilized the T-TESS during teacher candidates' clinical teaching experience. Again, this measure was used by school principals and faculty researchers, following the rubric which measured observations in a five-point Likert scale, ranging from Improvement Needed to Distinguished. There are 15 statements in this T-TESS, measuring the effectiveness of teachers in four domains (i.e., planning, instruction, learning environment, and professional practices and responsibilities). Results showed that the T-TESS average scores rated by university supervisors ranged from 1.80 to 3.00, with $M = 2.52, SD = .34$. The average scores rated by

faculty researchers ranged from 4.00 to 4.33, with $M = 4.18, SD = .14$. The average scores rated by school principals ranged from 2.22 to 3.67, with $M = 3.26, SD = .41$.

Non-parametric tests showed that there were significant differences in the T-TESS average scores rated by university supervisors, school principals, and faculty researchers. Overall, supervisors rated significantly lower than school principals and faculty researchers in 12 statements. School principals rated significantly higher than faculty researchers on the following two statements: 1) Data and Assessment Dimension 1.2, with $\chi^2(2) = 2.750, p = 0.018$, and 2) Professional Development Dimension 4.3, $\chi^2(2) = 2.691, p = 0.021$. University supervisors rated significantly lower than faculty researchers on the following 12 statements (see Table 2).

Table 2
T-TESS Results Comparisons

T-TESS	χ^2	p
Standards and Alignment Dimension 1.1	-2.799	0.015
Data and Assessment Dimension 1.2	-3.763	0.001
Knowledge of Students Dimension 1.3	-3.484	0.001
Activities Dimension 1.4	-3.042	0.007
Achieving Expectations Dimension 2.1	-2.416	0.047
Content Knowledge and Expertise Dimension 2.2	-3.377	0.002
Communication Dimension 2.3	-2.609	0.027
Differentiation Dimension 2.4	-2.753	0.018
Classroom Environment, Routines and Procedures Dimension 3.1	-2.671	0.023
Professional Demeanor and Ethics Dimension 4.1	-2.915	0.011
Goal Setting Dimension 4.2	-2.419	0.047
Professional Development Dimension 4.3	-3.112	0.006

Areas To Be Improved

Our findings indicated that candidates felt a lack of preparation in statement of understanding and use of multiple methods of assessment to engage learners in their own growth, monitor learner progress, and guide the teacher's and learner's decision making. Mentor cooperating teachers and principals also rated this statement with the lowest scores indicating the weakness of our program in preparing teacher candidates with adequate knowledge and skills in assessment.

According to principals, another aspect for program improvement is to develop candidates' leadership skills. For example, regarding a new teacher on her campus, the principal stated that "I would like to see her take on more of a leadership role and exposure to other grade levels. I am planning to utilize her expertise in preparing her to facilitate staff development."

Although data from principals, cooperating mentor teachers, and university supervisors did not indicate classroom management to be a weakness, during their interviews, beginning teachers spoke to classroom management skills as an area of growth. This would be an important area to consider, especially since many new teachers do not feel prepared in this area. As one new teacher stated:

"Across the board, it's discipline, especially the students that I have. I have a lot of helicopter parents and a lot of very expressive and verbally liberal parents. So, it's really on how to strategize on dealing with that and how to deal with it in the classroom. Another teacher discussed that he would like to learn more "ways to be a little more firm without being mean or shouting. I want to be effective, and, at the same time, there are elements where I feel like I can just get lax on, so I think classroom behavior and classroom management and not letting there be big gaps in between, that's something that I need to work on, personally."

Discussion

This study used multiple resources to evaluate our program by surveying candidates, graduates, mentor cooperating teachers, university faculty, and school principals, which was recommended by many researchers

(Bastian et al., 2018; Henry et al., 2012; Ruben, 2010). The results showed that our candidates were well prepared with an understanding of individual differences and the diverse backgrounds of learners. They were well prepared to create environments that support individual learning, collaborative learning, and encourage positive social interaction and active engagement in learning. This finding is contrary to the findings of a qualitative case study conducted by Lehman (2016, as cited in Lehman, 2017) that preservice teachers were lacking multicultural competence and needed professional development training. It is also in contrast with another case study (Rizzuto, 2017) who discovered early career teachers who taught English language learners (ELLs) lacked skills in working with diverse students and were reluctant to use a culturally responsive approach to teaching their ELLs. Our finding indicated the strength of our teacher candidates in terms of working with diverse students inside and outside of our teacher education program, which might be due to high-quality training with diverse students in the field.

Assessment seems to be a weak area that needs to be improved in our teacher education EC-6 and 4-8 program. Our findings revealed the weakness in preparing candidates with adequate knowledge and skills in assessing individual learners to guide instructions. Weakness in assessment to guide instruction seemed to be common and was also discovered in a national sample of beginning teachers (Darling-Hammond et al., 2002, as cited in Darling-Hammond, 2006). However, given limited studies focusing on developing assessment knowledge and skills among preservice teachers, it might be meaningful to research and develop additional curriculum that can be adopted during the teacher preparation program to focus on assessment and planning instructions based on assessment results. Professional learning communities formed by schools and districts can foster teachers' skills in the use of assessment to guide instruction (Hamilton et al., 2009). It is suggested by case study work (U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, 2010) that teachers' skills in assessment could be developed through their continuous practices on using students' data for instruction with facilitations from colleagues and mentor teachers.

Leadership skills should also be developed and practiced among our teacher candidates. Leadership skills are critical for school reform and improvement (Kise &

Russell, 2008, as cited in Xu & Patmore, 2012), which, ultimately, lead to enhanced student achievement (Gabriel, 2005, as cited in Xu & Patmore, 2012). However, less attention has been paid to preparing teacher leaders by providing a course on leadership skills in teacher education programs (Xu & Patmore, 2012). Teacher candidates lacking leadership skills seems to be a common problem not only in our teacher education program but also in others as well. For example, Silvernail (1998, as cited in Darling-Hammond et al., 2002) surveyed approximately 3000 beginning teachers, and results suggested that teachers who graduated from TPP received high ratings on only two items out of the four regarding instructional leadership. Formal training, such as attending a course (Xu & Patmore, 2012) and teacher-led professional development (Semadeni, 2010; Xu & Patmore, 2012), could be adopted to enhance leadership skills and should be encouraged in any teacher education program.

Limitations and Future Research

There are limitations related to this case study that could be addressed in future research. First is the small sample size of participants in this study, which might not provide a true picture of our TPP. The second limitation is the missing representatives of our teacher candidates from the early childhood grade levels, such as the ones who served at the childcare facilities. The findings could not be generalized to the early childhood TPP. A future research agenda would be to enlarge the sample size to include more participants. It may also be valuable to include not only more teacher candidates and school principals in different grade levels but also field supervisors, teacher peers who worked at the same school, and university faculty who taught our candidates to gain a more comprehensive perspective on the performance of our teacher candidates. The students taught by our beginning teachers could also be interviewed to understand their perceptions regarding teacher proficiency.

Conclusions

Despite the limitations, our study is significant in two aspects. First, we used multiple evaluation ratings to assess the effectiveness of our teacher candidate preparation and performance in their early career. Our data were comprehensive in scope, therefore, provided clearer directions for program improvement. Second, this study provided strong evidence of the benefits of using evaluating ratings in assessing TPP performance and effectiveness, which is echoed in research findings from Bastian et al. (2018). Especially when synthesizing all the feedback from multiple persons from clinical teaching to an early teaching career, evaluating ratings provides important information about teacher quality and clearer directions for what to improve and how to accomplish the improvement, which is currently limited in literature (Noell et al., 2019). Our study provided a good example of the utility of multiple performance evaluation ratings from different personnel inside and outside of TPP, which can be beneficial for both accountability and improvement of TPP.

Through this study, we gained a better understanding of our TPP in terms of strengths and areas of growth. Based on the results, our TPP should consider providing curriculum revisions to strengthen the teacher candidates in terms of their skills in assessment, classroom management, and leadership throughout our program. We will continue to monitor our teacher education program using multiple, comprehensive measurements to ensure our TPP continues to grow and improve.

References

- American Education Research Association. (2015). AERA statement on the use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452. <https://doi.org/10.3102/0013189X15618385>
- Bartell, T., Floden, R.E., & Richmond, G. (2018). What data and measures should inform teacher preparation? Reclaiming accountability. *Journal of Teacher Education*, 69(5), 426-428. <https://doi.org/10.1177%2F0022487118797326>
- Bastian, K. C., Patterson, K. M., & Pan, Y. (2018). Evaluating teacher preparation programs with teacher evaluation ratings: Implications for program accountability and improvement. *Journal of Teacher Education*, 69(5), 429-447. <http://doi:10.1177/0022487117718182>
- Bransford, J., Darling-Hammond, L., & Lepage, P. (2005). Introduction. In L. Darling-Hammond, & J. Bransford (Eds.), *Preparing teachers for a changing world. What teachers should learn and be able to do* (pp. 1- 39). Jossey-Bass.
- Coggschall, J. G., Bivona, L., & Reschly, D. J. (2012, August). *Evaluating the effectiveness of teacher preparation programs for support and accountability*. National Comprehensive Center for Teacher Quality. <https://eric.ed.gov/?id=ED543773>
- Council for the Accreditation of Educator Preparation. (2020). *Standard 4: Program impact*. <http://www.ncate.org/standards/2013/standard-4>
- Council for the Accreditation of Educator Preparation. (2020). *Standard 5: Provider quality, continuous improvement, and capacity*. <http://www.ncate.org/standards/2013/standard-5>
- Council of Chief State School Officers. (2013, March 1). *Interstate teacher assessment and support consortium InTASC model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development*. <https://ccsso.org/resource-library/intasc-model-core-teaching-standards-and-learning-progressions-teachers-10>
- Crowe, E. (2010, July 28). *Measuring what matters: A stronger accountability model for teacher education*. Center for American Progress. <https://fordhaminstitute.org/national/commentary/measuring-what-matters-stronger-accountability-model-teacher-education>
- Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, 57(2), 120-138. <https://doi.org/10.1177%2F0022487105283796>
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, 53(4), 286-302. <https://doi.org/10.1177%2F0022487102053004002>
- 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(2), 96-104. <https://doi.org/10.3102%2F0013189X15575031>
- Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <https://ies.ed.gov/ncee/wwc/practiceguide/12>
- Henry, G. T., Kershaw, D. C., Zulli, R. A., & Smith, A. A. (2012). Incorporating teacher effectiveness into teacher preparation program evaluation. *Journal of Teacher Education*, 63(5), 335-355. <https://doi.org/10.1177%2F0022487112454437>
- Lehman, C. (2017). Multicultural competence: A literature review supporting focused training for preservice teachers teaching diverse students. *Journal of Education and Practice*, 8(10), 109-116. <https://eric.ed.gov/?id=EJ1139702>
- Ludlow, L. H., Pedulla, J., Reagan, E., Enterline, S., Cannady, M., & Chappe, S. (2011). Design and implementation issues in longitudinal research. *Educational Policy Analysis Archives*, 19(11). <https://epaa.asu.edu/ojs/article/view/802>
- Monk, D. H. (2015). Reflections on teacher preparation. *Society*, 52(3), 219-224. <https://doi.org/10.1007/s12115-015-9889-z>
- Noell, G. H., Burns, J. M., & Gansle, K. A. (2019). Linking student achievement to teacher preparation: Emergent challenges in implementing value added assessment. *Journal of Teacher Education*, 70(2), 128-138. <http://dx.doi:10.1177/0022487118800708>
- Rizutto, K. C. (2017). Teachers' perceptions of ELL students: Do their attitudes shape their instruction? *The Teacher Educator*, 52(3), 182-202. <https://doi.org/10.1080/08878730.2017.1296912>
- Ruben, B. R. (2010). *Excellence in higher education guide. An integrated approach to assessment, planning, and improvement in colleges and universities*. National Association of College and University Business Officers.
- Semadeni, J. (2010, May 10). When teachers drive their learning. *Educational Leadership*, 67(8), 66-69. <http://www.ascd.org/publications/educational-leadership/may10/vol67/num08/When-Teachers-Drive-Their-Learning.aspx>
- Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
- Texas Education Agency. (2016, August 4). *Teacher handbook*. https://teachfortexas.org/Resource_Files/Guides/T-TESS_Teacher_Handbook.pdf
- Texas Education Agency. (2020). *STAAR resources*. <https://tea.texas.gov/student-assessment/testing/staar/staar-resources>
- U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2010). *Use of education data at the local level: From accountability to instructional improvement*. Author. Retrieved from <https://www2.ed.gov/rschstat/eval/tech/use-of-education-data/index.html>
- Xu, Y. J., & Patmor, G. (2012). Fostering leadership skills in preservice teachers. *International Journal of Teaching and Learning in Higher Education*, 24(2), 252-256. <https://www.isetl.org/ijtlhe/>

Zeichner, K. M., & Conklin, H. G. (2005). Teacher education programs. In M. Cochran-Smith, & K. M. Zeichner (Eds.), *Studying teacher education* (pp. 645-735). Lawrence Erlbaum Associates.