Does Infused ESOL Teacher Preparation Program Hold Promise Toward Narrowing the English Learner Achievement Gap? An Analysis of the One-Plus Model

***Nirmal Ghimire*[1]**
***Kerry L. Purmensky*[1]**
***Sushila Regmi*[2]**
[1]Department of Modern Languages and Literatures,
University of Central Florida
[2]Department of Social Work, Child Welfare and Social Policy,
Oslo Metropolitan University

Abstract

This study compared pretest and posttest scores of *n* = 5,469 K–12 students who were taught by *n* = 236 preservice teachers (PSTs) to assess if the infused ESOL One-Plus teacher preparation program (TPP) adequately trained PSTs to work with English Learners (ELs) and bridge the achievement gap. Results suggested that students had variable degrees of achievement strongly tied to their EL status, socioeconomic status, and the class size. There were increased learning gains among all students, a smaller achievement gap between student groups, and comparatively higher learning gains among EL, low-SES, and minority students compared to their counterparts after taught by the One-Plus PSTs.

Keywords: teacher preparation, achievement gap, English learners, ESOL infusion, teacher effectiveness, hierarchical linear modeling (HLM)

Introduction

Diversity in the mainstream U.S. classroom has been expanding rapidly for the last two decades, which necessitates greater accountability on the part of classroom teachers. As Darling-Hammond puts it, "the wide range of learning situations posed by contemporary students—who represent many distinct languages, cultural, and learning approaches"—require "a much deeper knowledge base about teaching for diverse learners than ever before" (Darling-Hammond, 2008, p. 304). However, research findings suggest that most in-service teachers lack the necessary attitude, disposition, and professional skills to create inclusive classrooms (van Laarhoven et al., 2007; Zinth, 2013).

Nirmal Ghimire is Postdoctoral Fellow at K–16 Literacy Center in the School of Education, College of Education and Psychology at the University of Texas at Tyler.

Kerry L. Purmensky is a Professor in the Department of Modern Languages and Literatures, and the Director of TESOL Program at the University of Central Florida, Orlando.

Sushila Regmi completed her Master's in International Social Welfare and Health Policies from Oslo Metropolitan University, Oslo, Norway.

Students' linguistic background and ethnicity are the two most common indicators of the achievement gap in the United States (Lancour & Tissington, 2011; White, 1982). English Learners (ELs) routinely score below average on standardized tests and are more likely to be taught by undertrained teachers (Gándara & Maxwell-Jolly, 2006). Historically, ELs in U.S. schools were either in "English Only" (e.g., Sheltered English, Structured English Immersion), or "pull-out ESL" classrooms with generalist teachers, who were not properly trained to work with ELs (Barrow & Markman-Pithers, 2016). They were either "pulled out" from regular classrooms to participate in English as a Second Language (ESL) classes or they took ESL instructions during a regular class period as a separate standalone course like math and science. In either case, ELs were taught by ESOL professionals with specialized language and pedagogical skills. Despite many benefits, such programs over the decades established a sense of exemption from responsibility on behalf of generalist teachers as they perceived ELs as the purview of ESOL specialists (Coady et al., 2015; McHatton & Parker, 2013).

In a similar vein, researchers suggest that classroom communication gaps (Nutta et al., 2012), linguistically demanding tests and classes (Abedi, 2002), traditional teacher education programs that do not account for the changing paradigm (Yoder et al., 2016), an unwelcoming school and classroom environment (Berkowitz et al., 2016), and a lack of systematically articulated district EL plans (Good et al., 2010) are some of the additional reasons for such disparity.

Achievement gaps can accelerate negative reactions in school communities, such as segregating students (Welsh, 2018), promoting the use of remedial curricula (Beecher & Sweeny, 2008), perpetuating stereotypes that affect low-income, low-achieving students (Borman et al., 2005; Furgione et al., 2018), and fostering inequitable treatment (Carrell & Dittrich, 1978).

The infused ESOL One-Plus model (One-Plus model henceforth) in teacher preparation is one of many programs that take aim at the existing EL achievement gap and strive to utilize research-backed components within the teacher education curricula (Nutta, et al., 2012). The One-Plus model helps preservice teachers (PSTs) master the theoretical underpinnings of comprehensible English language input and have firsthand experience of implementing it in the classrooms with ELs, repeatedly. PSTs are taught to create meaningful oral inputs for their ELs besides being consistently exposed to specifically designed course works and projects to enhance their cultural and linguistic responsiveness (Lavery et al., 2018). To address the 21st century classroom needs posed by diverse student groups including ELs, the One-Plus model implements enhanced curriculum, extended service learning and internship requirements, and a cutting-edge simulation lab offering a virtual environment for PSTs to interact with ELs of varying English proficiencies (Nutta et al., 2012).

As One-Plus is a new teacher preparation model, there is a lack of research on its effectiveness in narrowing EL gaps, the sole essence of this program. On the other hand, compared to the amount of debate over the achievement gap in the last few decades, the volume of quantitative research about the achievement of ELs is thin (Jaynes, 2015). Using archival Teacher Work Sample (TWS) data from Fall 2016 and Spring 2017, this study attempts to assess whether the One-Plus model prepared PSTs to narrow the EL achievement gap and answer the following research questions:

1. Were One-Plus PSTs able to narrow the EL achievement gap in their internship classrooms?
2. Did the rate of change in students' posttest scores vary by the One-Plus PSTs' major and the grade level that they taught?

Infused ESOL One-Plus Model

Language fluency is directly associated with ELs' accomplishment in content area courses (Anthony & Walshaw, 2009; McClain & Cobb, 2004), because it helps eliminate confusion and challenges while scaffolding prior knowledge in the everyday classroom (Lavy & Mashiach-Eizenberg, 2017). The importance of English language teaching skill among classroom teachers is not only well understood but also encouraged. It is believed that by making classrooms linguistically accessible to all students —learned through quality teacher preparation curricula—teachers can accelerate learning outcomes among ELs (Hunter & Bartee, 2003; Babinski et al., 2018).

Recent development in teacher preparation programs highlights the importance of providing specified training for generalist teachers and administrative personnel in line with the recommended best practices for ELs (Wheeler & Govoni, 2014). In the state of Florida, a Consent Decree between the State Board of Education and the League of United Latin American Citizens (LULAC) paved the way toward the milestone infusion model that mandated teacher preparation programs to infuse five key ESOL domains in teacher preparation curricula (Nutta et al., 2012; Wheeler & Govoni, 2014):

- cross-cultural communication and understanding
- language and literacy
- methods of teaching ELs
- ESOL curriculum and material development
- ESOL testing and evaluation

The Infusion model maintains that PSTs apply and demonstrate their acquired knowledge in these domains upon successful completion of their programs. It also highlights the need for field placements in classrooms with ELs for the mastery of learned knowledge through coursework (Conderman et al., 2012; Gehrke & Cocchiarella, 2013; van Laarhoven, et al., 2007).

Nutta et al., (2012) defines ESOL Infusion as an "approach to ensuring that teacher candidates are prepared to teach and assess ELs in a mainstream classroom," which involves "integrating, rather than simply appending curricular content and assessment that promotes the development of the desired knowledge, skills, and disposition" (p. 20). Furthermore, the EL content in teacher preparation courses should comprise "EL-focused topics, objectives, instructional materials and media, in-class activities, course assignments, field/clinical experiences, and assessments" and should be built into "EL-specific courses" and distributed across all teacher preparation courses (Nutta et al., 2012, p. 20).

Promoting three critical qualities of interconnectedness, cohesion, and interdisciplinarity, the One-Plus model endorses comprehensive, curriculum-wide infusion by ESOL faculty and specialists in collaboration with content area faculty, administrators, and other stakeholders (Nutta et al., 2012). The One-Plus model claims to be flexible in terms of course requirements, i.e., one EL-specific course for all programs leading to EL-qualified credentials (EL-qualified for academic subjects like math or science); two for English language arts majors (EL-qualified for language arts); and more than two EL-embedded and EL-infused courses for the PSTs seeking ESOL endorsement or certification. It also includes a "basic coverage" option for teacher preparation programs (TPPs) that are in the states where ESOL endorsement is not a requirement or if a program is unable to add coursework to their existing curricula. Basic coverage, at the minimum, consists of ESOL embedded foundational and content area courses, including a reflective portfolio of PST compiled EL-focused assignments (Nutta et al., 2012).

Effectiveness of the Infusion Model

Given that the One-Plus model lacks research-based evidence to make conclusive remarks on the model and its effectiveness (Boyd et al., 2009; Nutta et al., 2015), the preliminary findings, especially from individual case studies and small group qualitative and quantitative studies, demonstrate that it outperforms the traditional TPPs (Conderman et al., 2012; Smith, 2011).

One of the previous works that directly measured One-Plus PSTs' effectiveness was a doctoral dissertation, Ghimire (2020). This study also tapped into its trends over a period of five semesters using the pretest and posttest scores of (n = 20,809) K–12 students who attended classes taught by (n = 768) One-Plus PSTs from 2016 through 2018. In addition, this study also measured the moderation effect of PSTs' area of study, class size, and grade levels taught (Level 2 variables) on ELs' test scores. The findings outlined wide achievement gaps among student groups clearly aligned to their sociodemographic characteristics (Ghimire, 2020). The highest gap, i.e., approximately 9.7 units (percentage points), was between EL and non-ELs in which roughly 9% of the variability in students' pretest scores was accounted for by their EL status. Furthermore, the study found lower test scores among the low-SES ELs, and ELs with exceptionalities, compared with the high-SES ELs and ELs without any exceptionalities. The baseline achievement gaps as measured by the pretest models shrunk by approximately 50% between EL and non-EL students, by 40% between low and high-SES students, and 38% between students with and without exceptionalities, respectively. Overall, the study found a positive linear growth in students' achievement scores over the period of five semesters (Ghimire, 2020).

Similarly, Lavery et al. (2018) compared pretest and posttest scores of 8,326 students enrolled in 288 One-Plus PSTs classrooms ($n = 288$) to gauge if the PSTs received enough preparation to support EL achievement in mainstream K–12 classrooms. In line with the previous studies on this topic, the results showed that ELs had lower classroom pre- and post-assessment scores than non-ELs. The study also found that the difference of marginal estimated means between EL and non-ELs were statistically significant, and the student-level variables significantly explained their post-assessment scores. Among the student-level variables, disability and EL status were the strongest predictors of higher achievement gaps (Lavery et al., 2018).

A similar study was conducted by Gherke & Cocchiarella (2013), which surveyed 125 PSTs enrolled in secondary education, special education/dual certification, and elementary education programs with varying degrees of EL focused courses, field experiences, and student teaching requirements. The results showed that the PSTs' ability to identify characteristics of efficient classroom inclusion was directly tied to the components of TPPs, i.e., PSTs who had to go through the most extensive EL requirements were found to be more capable of identifying such elements.

After conducting a multiple analysis of variance (MANOVA) on a set of survey data collected from 273 Infusion PSTs, Smith (2011) found that candidates who took one ESOL-specific/embedded course during their teacher preparation had statistically significantly lower perception of their knowledge and skills on the pre-course measures, while those perceptions remained comparable during the post-course measures. Likewise, collecting data from 31 New York TPPs, approximately 36,000 students, and the numerous TPP faculties, Boyd et al. (2009) discovered that (a) on average, TPPs that produced teachers who were able to increase student learning prepared PSTs who were more effective in EL teaching; and (b) the TPPs that gave higher importance to the portfolio of work done in classroom with students and field experiences had statistically significantly higher student learning outcomes than the ones that did not. The findings of these studies indicate that the TPPs that prioritize the amount and volume of EL-focused courses, activities, teaching, and field experiences are better equipped to narrow the EL achievement gap.

In addition, there are a few small-scale studies that measured the effectiveness of either the infused ESOL TPPs or some aspects of effective practices recognized by the infusion model. Some of them also reported teacher educators' firsthand experiences while completing infused ESOL professional learning activities (e.g., Costa et al., 2005; de Oliveira & Athanases, 2007; Hutchinson, 2011; Isaac & Quantz, 2011; Meskill, 2005; Niday, 2012). Findings of these studies strengthen what van Laarhoven et al. (2007) and Conderman et al. (2012) concluded: Teacher candidates who attend TPPs with higher requirements for EL infused courses, course-specific competencies, and/or clinical experiences are associated with more substantial gains in students' test scores. Based on the findings of these studies, the infused ESOL One-Plus model seems to incorporate required curricular, instructional, and assessment skills in TPPs to foster comparable learning gains between EL and non-ELs.

Methods

This non-experimental correlational research was conducted using archival Teacher Work Sample (TWS) data available through the institutions' LiveText portal. The data were reported by One-Plus PSTs after successful completion of a semester long internship during fall 2016 and spring 2017. There was a total of 314 PSTs who taught 6,920 students (all grades) in schools located in the school districts around the institution where the One-Plus model is implemented. Eight One-Plus PSTs failed to report the program of study, 25 of them did not provide the demographic information, and 14 of the data files did not have pretest or posttest scores. The researchers cross referenced the sample and discovered that 31 teacher candidates did not report multiple student-level indicators (e.g., EL status, grade levels, FRPL status, etc.), which brought the final dataset of $N = 236$ One-Plus PSTs and their $N = 5{,}561$ students.

The TWS is a widely used instrument that records the PSTs' capability to design, implement, and assess standards-based student learning and reflection on the learning process (Denner et al., 2001; Foster et al., 2010). At the institution of study, the TWS is a two-tier process: (a) prompt section, which records One-Plus PSTs' subjective account of their ability to locate and gather information, such as critical thinking, reasoning, analyzing etc. (Benton et al., 2012), and (b) GraphMaker™, which is an Excel based spreadsheet that captures students' demographic characteristics and test scores. This research was conducted using the GraphMaker™ data. Table 1 lists the variables used in this study and their descriptions.

Table 1. *Measures: Variable Descriptions*

Variables	Descriptions
Level 1, Student-Level Variables:	
EL	A dichotomous predictor representing students' English language status. Students who speak English as their first language were non-ELs (coded 0) and ELs who speak English as a second language (coded 1).
FRPL	A dichotomous predictor denoting students eligible for free or reduced-price lunch program (low-SES) coded 1, otherwise coded 0.
ETHNICITY	A nominal predictor derived from students' race. non-Hispanic Whites were coded 0, Blacks were coded 1, and Hispanics, Asian Pacific Islanders, Alaskan Natives or American Indians, & other or multiracial students were coded 2, 3, 4, 5, respectively. Furthermore, the researchers dichotomized this variable and named the new variable (WHITE), in which, the non-Hispanic White were coded 0 and 1 otherwise.
GENDER	A dichotomous student-level predictor denoting students' gender. Male students were coded 1 and females 0.
PRETEST	Both a predictor and outcome variable. It was used as the outcome variable in the pretest model that assessed baseline achievement gaps. It was also used as a predictor for the

	posttest model to avoid the spurious regression result common among the nonstandard test scores.
POSTTEST	A continuous percentage score the students obtained in their posttests showing students' learning gains.
Level 2, PST-Level Variables:	The Level-2 variables were not directly evaluated for their main effect on students' test scores but checked for their moderation effects.
STUDENT	A continuous variable resembling the total number of students per class per PST.
TPROGRAM	A categorical variable, where PSTs from elementary education (ELEM) were coded 0, math education (MATH) coded 1, and English language arts (ELAE) were coded 2.
TGRADE	Originally, it was a discrete count variable denoting the grade levels the PSTs taught. However, it was changed to a categorical variable with three independent groups, i.e., elementary school (K– Grade 6), middle school (Grades 7 – 8), and high school (Grades 9 – 12). They were coded 0, 1, 2, respectively.

Analytical Strategies

A two-level hierarchical linear model (HLM) was used to measure the main effect of the Level-1 predictors, and the moderation effects of the Level-2 variables in assessing EL and non-EL achievement gaps. In general, hierarchical designs assess the impact of the interdependence of the control variable including interactional effects with Level-2 covariates (Raudenbush & Bryk, 2002). To facilitate the estimation and interpretation of the student-level variables, they were group-mean-centered, and the teacher-level variables were grand-mean-centered. Appropriate to most two-level analyses, the researcher used full information maximum likelihood (FIML) estimation to account for the desired number of parameters in the final model (McCoach, 2010).

The statistical analysis was conducted using HLM v.8 and Statistical Package for the Social Sciences (SPSS) software. Descriptive statistics are reported in Table 1. As seen, slightly more than 10% of the total students were ELs, while approximately 51% of students were eligible for free or reduced-price lunch. Students were fairly distributed based on their gender. Approximately 46% of the students were non-Hispanic White followed by the Hispanic population, which comprised 31% of the entire sample. The research area had higher concentrations of the Hispanic population compared to the national average of 18.9% in the year 2020 (U.S. Census Bureau).

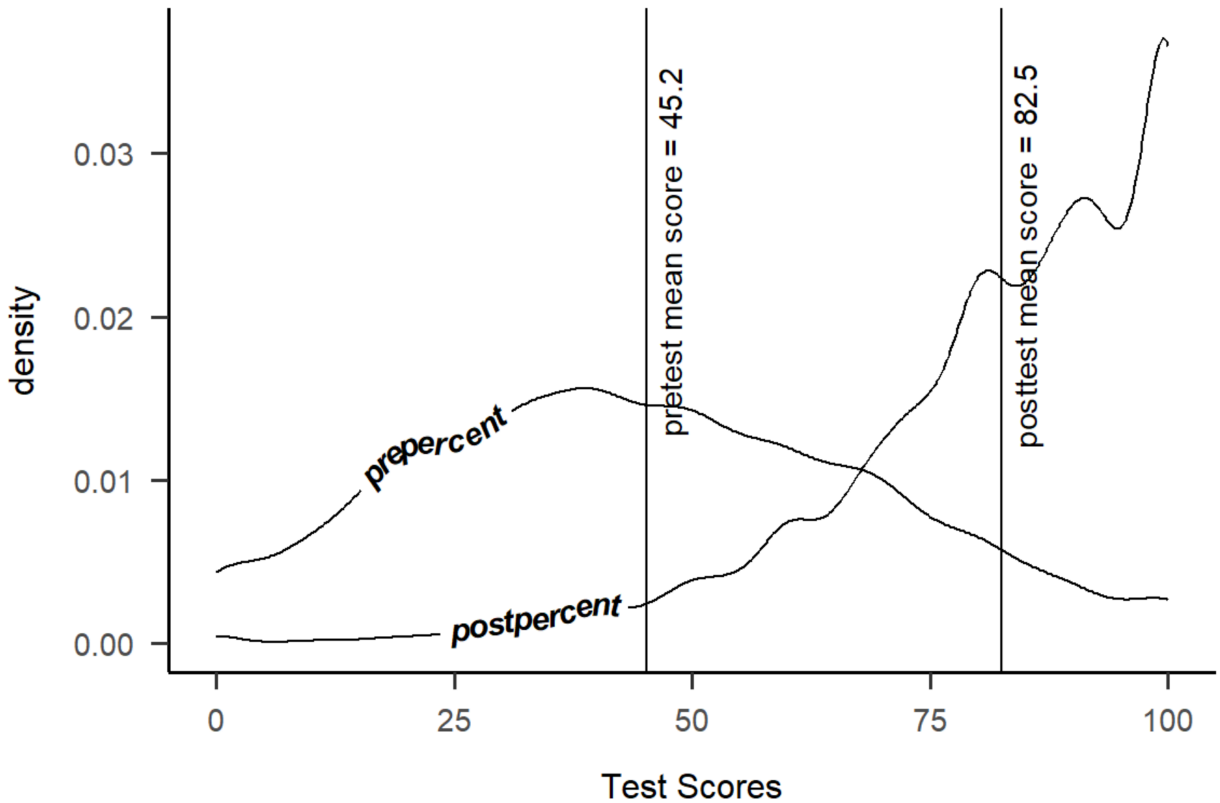
Table 2. *Descriptive Statistics of the Variables Used in the Models*

Variables	Student		Teacher	
	n	%	*n*	%
Level 1, Student-Level Predictors				
FRLUNCH				
Yes	2,837	51.02		
No	2,724	48.98		
EL				
ELs	583	10.48		
Non-ELs	4,977	89.50		
MALE				
Male	2,839	51.05		
Female	2,721	48.93		
ETHNICITY				
Non-Hispanic White	2,577	46.34		
Hispanic or Latino	1,704	30.64		
American Indian or Alaskan Native	16	0.29		
Non-Hispanic Black	798	14.35		
Asian or Pacific Islander	273	4.91		
Multiracial or not listed	192	3.45		
WHITE				
White	2,577	46.34		
Non-White, Minority students	2,984	53.66		
Level 2, PST-Level Variables				
TPROGRAM				
Elementary Education	4,627	83.20	220	93.22
Math Education	154	2.77	5	2.12
English Language Arts	780	14.03	11	4.66
TGRADE				
Elementary School	4,627	83.20	218	92.37
Middle School	324	5.83	7	2.96
High School	610	10.97	11	4.66
STUDENT	5,561			

Note. Preservice teachers $N = 236$. EL = students' English learner status; FRPL = students' free/reduced price lunch status; MALE = students' gender variable; ETHNICITY = students' ethnicity; WHITE = dichotomized ethnicity variable; TPROGRAM = preservice teachers' major; TGRADE = grade level the preservice teachers taught; PRETEST = students' pretest scores; POSTTEST = students' posttest scores.

The average student teacher ratio was 23.56. Overwhelming numbers of PSTs were elementary education majors who taught elementary grade students. The average pretest score was ($M = 45.20$, $SD = 24.16$), while the average posttest score was ($M = 82.50$, $SD = 17.12$). Figure 1 shows the comparative distribution of pretest and posttest scores for all the samples included in the study. The pretest scores are fairly distributed among the participants, while the posttest scores seem to be left skewed, suggesting growth on all students regardless of their sociodemographic statuses.

Figure 1. *Overall Distribution of Students' Pretest and Posttest Scores*



Model Building and Results

The analyses were conducted in three phases. First, a baseline model, i.e., unconditional/null model with fixed intercepts was estimated to measure the amount of variability in test scores among students (as a function of their sociodemographic characteristics) and between classrooms (as a function of PSTs' characteristics). The average pretest and posttest estimated means were 45.2, and 82.5 and they were statistically significantly different from zero. Statistically significant variations in the mean outcome existed in both cases as explained by [$\tau_{(PRETEST)} = 289.64$, $\tau_{(POSTTEST)} = 85.6$]. There was statistically significant change in log likelihood (*–2LL*) after allowing the intercept to vary randomly $\chi^2(1) = 2696$, $p < .0001$, letting us know that the intercept varies significantly across PSTs. The researchers calculated a series of intra-class correlations (ICC) to measure variation in test scores among student groups and denote change during the model building phase. For the pretest null model, the ICC was 0.4803, which was 0.2856 for the posttest null model. The findings suggested that approximately 48% of the variation in students' pretest scores and 29% of the variation in posttest scores were explained by Level 1 variables while 52% in pretest scores, and 69% in posttest score models remained unaccounted for. The variability in students' pretest scores based on their characteristics decreased to 47% in the final pretest model, while it increased to 30% in the final posttest model, which helped us conclude that the final models were better fitting models than the null models.

In the second stage, Level 1 variables were included in the model one at a time starting with fixed slopes at first and then allowing them to vary randomly to check if the more parameterized models were better fitting models. In addition, the researchers compared the Akaike information criteria (AIC) and Bayesian information criteria (BIC) among the pretest models along the way until the final pretest Level 1 model was identified. The Level 1 predictors without having statistically significant impacts on students' pretest scores were excluded from the model. Afterward, the researchers plugged in the Level 2 variables to check if they moderated any change in students' pretest scores.

The results of the Model 1 (null model), Model 2 (with Level 1 variables only), and Model 3 (with all Level 1 and Level 2 variables) are presented in Table 3. The Model 2 results showed that the average pretest scores remained the same, but it was statistically significantly better than the null model. The ELs had statistically significantly lower pretest scores $\beta_{EL(MODEL2)} = –7.60$, $t_{(235)} = –7.72$, $p < .001$, after controlling for their FRPL status, gender, and ethnicity. This gap increased to –8.10 after including PST moderators in the model. Male students had statistically significantly higher pretest scores in both Models 2 and 3 compared to the female students. Unlike many prior studies, this study did not find any statistically significant differences in students' test scores based on their ethnicity in both Model 2 ($\beta_{ETHNICITY(MODEL2)} = –0.23$, $SE = 0.17$) and Model-3 ($\beta_{ETHNICITY(MODEL3)} = –0.25$, $SE = 0.17$). As the ethnic gaps were not in the center of this study, the researchers did not run the post-hoc analyses of this variable, rather dichotomized it and studied differences in the test scores between minority students and their White counterparts.

Likewise, the gap between low-SES students ($M = 41.11.50$, $SD = 23.80$) and their counterparts was the second largest in both Model 2 ($\beta_{FRPL(MODEL2)} = –5.28$, $SE = 0.78$) and Model 3 ($\beta_{FRPL(MODEL3)} = –5.56$, $SE = 0.78$). The PST variables did not have statistically significant moderation effects on students' pretest scores. Increase in one grade level unit, i.e., from elementary ($M = 46.76$, $SD = 24.66$) to middle ($M = 29.93$, $SD = 24.70$), or middle to high school ($M = 40.96$, $SD = 16.86$) was associated with a 2.16 unit drop in students' pretest scores, and this decrease was statistically significantly higher than zero. The grade level the PSTs taught (*TGRADE*) and their area of study (*TPROGRAM*) statistically significantly moderated male and female students' pretest scores. Compared to female students, a unit increase in a grade level accounted for approximately 1 unit decrease in male students' pretest scores as evident in ($\beta_{GENDER*TRGRADE} = –0.93$, $SE = 0.28$).

Table 3. *Pretest Model Comparison*

Variables	Baseline Model	Model 2	Model 3
English Learner Status	–	–7.60***	–8.09***
(ref = non-English Learners)	–	(0.98)	(0.97)
Students' Gender	–	1.51***	1.57**
(ref = Female)	–	(0.48)	(0.50)
Gender*TPROGRAM	–	–	4.64***
	–	–	(1.24)
Gender*TGRADES	–	–	–0.93***
	–	–	(0.28)
Students' Ethnicity	–	–0.23~	–0.25~
(ref = non-Hispanic White)	–	(0.17)	(0.17)
Free-Reduced Price Lunch Status	–	–5.28***	–5.56***
(ref = non-Eligible aka. high-SES students)	–	(0.78)	(0.78)
Intercept (Pretest Score)	45.20***	47.18***	47.23***
	(0.33)	(1.14)	(1.09)
Intercept *TGRADES	–	–	–2.16***
	–	–	(0.45)
N Preservice Teachers	236	236	236
N Students	5,561	5,561	5,561

Note. Analyses weighted using pretest scores; standard error in parentheses; TGRADES = grade level the preservice teachers taught; TPROGRAMS = preservice teachers' major
~p >.10. *p < .05. **p < .01. ***p < .001.

Final Models, Assumption Testing, and the Results

The final phase included the final analyses after identification of the best pretest model which was then extended to the posttest model for meaningful comparison of teaching effectiveness. The final pretest and posttest structure models are expressed herewith:

Pretest Level 1 Model

$$PRETEST_{ij} = \beta_{0j} + \beta_{1j}(MALE_{ij}) + \beta_{2j}(FRLUNCH_{ij}) + \beta_{3j}(EL_{ij}) + \beta_{4j}(WHITE_{ij}) + r_{ij}$$

Posttest Level 1 Model

$$POSTTEST_{ij} = \beta_{0j} + \beta_{1j}(MALE_{ij}) + \beta_{2j}(FRLUNCH_{ij}) + \beta_{3j}(EL_{ij}) + \beta_{4j}(PRETEST_{ij}) + \beta_{5j}(WHITE_{ij}) + r_{ij}$$

Pretest & Posttest Level 2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(TPROGRAM_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{21}(TPROGRAM_j) + \gamma_{22}(TGRADE_j)$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + u_{4j}$$

Based on the testing of different models, total number of students per One-Plus PST (*STUDENT*) did not statistically significantly moderate the change in students' pretest scores, thus, this variable was excluded from the model above. The grade level the PSTs taught moderated the average pretest scores, and the pretest scores of the male and female students. The slope of the variables EL and WHITE varied randomly, while the FRLUNCH and MALE did not.

Upon deciding the final models, the researcher evaluated the data for basic assumptions. The test of homogeneity of Level 1 variance was statistically significant ($\chi^2 = 1581.06$, $df = 188$, $p < .001$) for pretest, and ($\chi^2 = 897.45$, $df = 188$, $p < .001$) for posttest, showing variability among Level 2 in terms of the residual of Level 1 variance. The chi-square statistics for pretest model was significant, χ^2 (163, $N = 236$) = 975.26, $p = 000$, for the pretest model indicating that the null hypothesis of homogeneity of the Level 1 variance was rejected. In other words, the results showed that there was variability among the ($J = 163$) Level 2 (i.e., PST-Level) units in terms of the residual within-students (i.e., Level 1) variance. Similarly, the chi-square statistics were significant χ^2 (162, $N = 236$) = 877.30, $p =$ 000, for the final posttest model indicating that the null hypothesis of homogeneity of the Level 1 variance was rejected. The histogram and normal Q-Q plot of Level 1 residual were normal. Moreover, residual analysis for Level 2 normality was conducted and the results showed that the normality and homoscedasticity were fair assumptions. The results of final pretest and posttest models are reported in Table 4.

Table 4. *Solutions for Fixed Effects of Pretest and Posttest Models.*

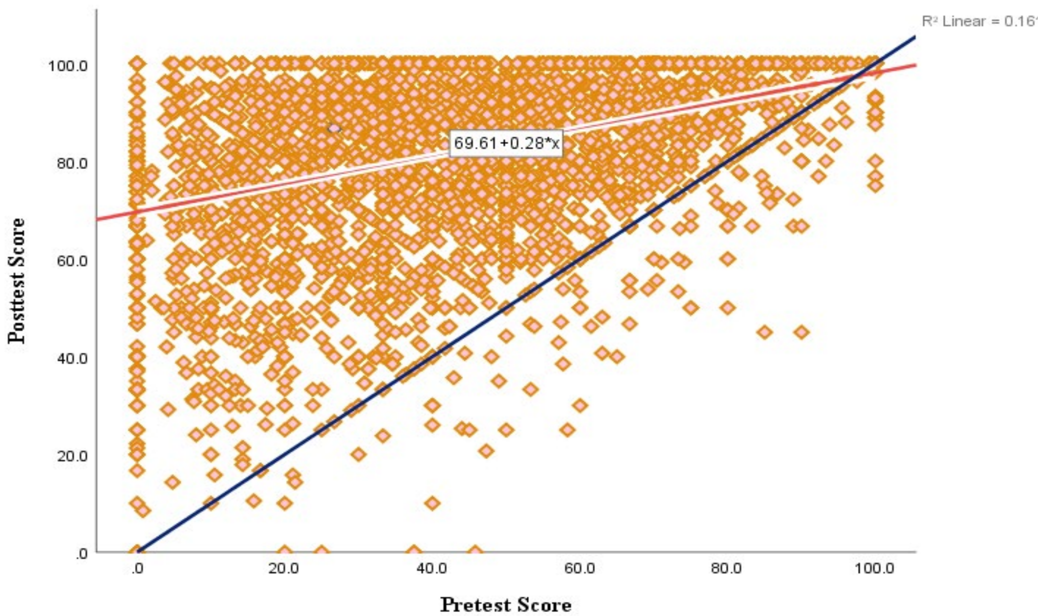
Variables	Final Pretest Model	Final Posttest Model
English learner status	–7.16***	3.98***
(ref = non-English learners)	(1.00)	(0.70)
Students' gender	1.56**	1.13***
(ref = Female)	(0.50)	(0.34)
Gender*TPROGRAM	4.85***	–0.68~
	(1.25)	(0.93)
Gender*TGRADES	–0.98***	–0.04~
	(0.28)	(0.21)
Students' minority status (WHITE)	2.26***	0.79*
(ref = Minority student population)	(0.57)	(0.37)
Free-reduced price lunch status (FRPL)	–4.45***	–0.87*
(ref = non-Eligible aka. high-SES students)	(0.60)	(0.45)
Pretest scores (as predictor)	–	0.33**
	–	(0.01)
Intercept	47.22***	82.52***
	(1.08)	(0.62)
Intercept*TGRADE	–2.12***	–1.17***
	(0.45)	(0.20)
N Preservice teachers	236	236
N Students	5,561	5,561

Note. Analyses weighted using pretest scores; standard errors in parentheses. TGRADES = Grade level the preservice teachers taught; TPROGRAMS = preservice teachers' major; ELEMENTA = if PSTs were in Elementary Education Track.
$\sim p > .10$. $*p < .05$. $**p < .01$. $***p < .001$.

On average, minority students had 2.26-unit lower pretest scores compared to non-Hispanic White students, and the gap was statistically significantly higher than zero. As evident in all pretest models, the gap between ELs and their non-ELs counterparts were the biggest, followed by the gaps between low- and high-SES students.

Likewise, the results of the posttest model showed that the average posttest score was, $\beta_{POSTTEST}$ = 82.52, t (234) = 132.86, $p < .001$, which was approximately 75% higher than the average pretest scores. Statistically significant gaps among most of the student groups still existed, however they were much smaller than the pretest models. The achievement gap between male and female students dropped by approximately 28%; the EL, non-EL gap by approx. 40%; minority and non-minority by approximately 65%; when the highest reduction rate (i.e., 80%) was observed between low- and high-SES students. Moreover, the moderation effect of the Level 2 variable TGRADE on the average test score was reduced by about 45%, while the moderation effect of grades the PSTs taught, ($\beta_{GENDER*TPROGRAM}$ = –0.68, SE = 0.93), and PSTs' program of study, ($\beta_{GENDER*TGRADE}$ = –0.04, SE = 0.21) were no longer statistically significant. In addition, students' pretest scores were a statistically significant predictor of their posttest scores, β = 0.33, t (235) = 23.80, $p < .001$. As can be seen in Figure 1, there was a linear positive relationship between pretest and posttest scores, i.e., one percentage point increase in students' pretest scores were associated with roughly 0.33-unit higher posttest scores.

Figure 2. *Scatter Plot Showing Fit Line for Posttest Score & Linear Relationship Between Pretest and Posttest Scores*



Effect sizes of the individual Level 1 predictors for the posttest model were measured in terms of Cohen's *d* (Cohen, 1988). Based on the results, gender had a small effect size, $d_{GENDER} = 0.11$; socioeconomic status, $d_{FRPL} = -0.29$, and minority status, $d_{WHITE} = -0.23$ had medium effects. The highest effect size was associated with students' EL status, $d_{EL} = -0.48$, but it was still a medium effect.

Randomly Selected Samples

The One-Plus PSTs in this study created their own pretest and posttest based on the learning goals they set. Thus, there was a lack of consistency in these tests. To check for the robustness of the findings, researchers broke the total sample size into two (60% and 40%) random subsamples using the "Select Cases" function in SPSS. The result of a single coin toss helped researchers decide which 60% random subsample to analyze. The 60% subsample contains all *N* = 236 One-Plus PSTs and their *N* = 3,305 students. Identical HLMs were conducted replicating the final pretest models. Of these samples, roughly 10.38% were ELs, 53% were identified as minority students, 49% received FRPL and 50.71% were males. The results of both 60% subsample pretest and posttest models are presented in Table 5.

Table 5. *Solutions for Fixed Effects of Pretest and Posttest Scores Analyses Using 60% Random Samples*

Variables	Final Pretest Model	Final Posttest Model
English learner status	–7.47***	–6.04***
(ref = non-English learners)	(1.27)	(1.01)
Students' gender	1.93**	1.25*
(ref = Female)	(0.66)	(0.48)
Gender*TPROGRAM	–2.32~	–1.35~
	(1.70)	(1.23)
Gender*TGRADES	–0.42~	0.20~
	(0.37)	(0.27)
Students' minority status (WHITE)	–2.19***	–0.09~
(ref = Minority student population)	(0.76)	(0.52)
Free-reduced price lunch status (FRPL)	–4.59***	–1.52*
(ref = non-Eligible aka. high-SES students)	(0.80)	(0.62)
Pretest scores (as predictor)	–	0.33**
	–	(0.02)
Intercept	47.10**	82.53***
	(1.09)	(0.64)
Intercept*TGRADE	–2.23***	–1.16***
	(0.46)	(0.22)
N Preservice teachers	236	226
N Students	3,305	3,305

Note. Analyses weighted using pretest scores; standard errors in parentheses. TGRADES = Grade level the preservice teachers taught; TPROGRAMS preservice teachers' major; ELEMENTA = if PSTs were in elementary education track.

~$p > .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

A comparative study across all runs showed that the sign and magnitude of most of the parameter estimates were similar to the results of the final pretest and posttest models presented in Tables 4 and 5. A notable change (i.e., a larger gap) was observed in the case of EL and non-ELs in the 60% subsample posttest results. The statistically significant gap between non-Hispanic White students and their minority counterparts in the final posttest model did not exist in the 60% random samples. The consistency of the findings between all other variables shows that students' posttest scores were better explained by students' gender, socioeconomic status, and pretest scores than by students' EL status.

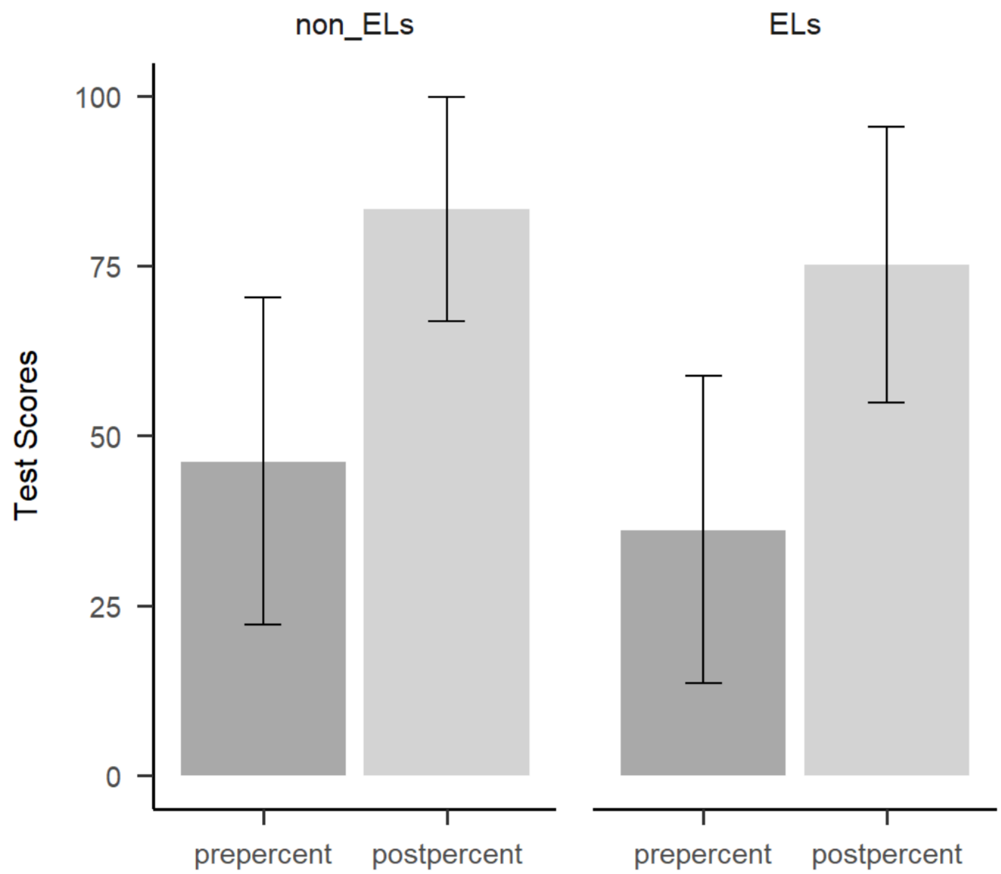
Discussion

This study found that the One-Plus PSTs were able to narrow EL-achievement gaps in their internship classrooms as measured by pretest-posttest data. The data indicated that the rate of change in students' posttest scores showed no variation based on the One-Plus PSTs' major. After controlling for students' EL or SES-status, gender, ethnicity, and PSTs' major, number of students per class, the grade levels the PSTs taught had negative impact on the students' test scores. The results of the posttest model showed that including targeted EL teaching skills in teacher preparation models could be instrumental in bridging EL achievement gaps.

Achievement Gap

We acknowledge the value that ELs bring to the classroom; however, the data indicated a persistent achievement gap at all grade levels. Based on the results, the achievement gap aligned to other sociodemographic characteristics such as ethnicity, SES, and sometimes gender, as well. As usual, students' EL status, as expected, was found to be associated with the greatest gap. Moreover, gaps existed in both pretest and posttest models, but the gaps in the posttest model were much smaller than in the pretest model. The bar plot in Figure 3 depicts students' pretest and posttest scores based on their EL and non-EL statuses.

Figure 3. *Framing Test Scores Based on Students' English Learner Status*



Overall, the pretest and posttest mean scores for students of different demographic characteristics were consistent with previous studies (Ghimire, 2020; Lavery et al., 2018). Based on the previous research, EL status interacts significantly with SES while measuring the achievement gap, and low-SES ELs tend to score lower than high-SES ELs (Ghimire, 2020; Kieffer, 2010). This study, however, did not find EL-status interacting statistically significantly with any other variable, including SES status. Further research needs to explore why these two factors, often the two greatest contributors to the achievement gap, did not influence each other in this study.

One-Plus Model Narrows the Achievement Gap

The data provided some evidence that the One-Plus PSTs were able to narrow the EL achievement gaps in their internship classrooms. Like the baseline model, the posttest model found statistically significant differences in EL academic achievement, however, the gap in posttest models were approximately 40% lower. This finding is consistent with prior research, e.g., Ghimire (2020), and

Lavery et al. (2018). While we did not investigate the specifics of the One-Plus model, the researchers speculate that the reason may be any of the service-learning with ELs, a simulation lab experience, enhanced EL-infused courses, and teaching differentiating instruction for ELs component or the interaction of all of these factors. Research focusing the individual aspects of the One-Plus model may shed more light on this.

There are some indicators that EL teaching techniques such as differentiating instruction, scaffolding, and grouping according to language ability have a positive impact on non-EL achievement (Goldenberg, 2013). Coursework in the One-Plus model focuses on preparing PSTs to recognize, understand, and address the needs of ELs at different levels of language ability (Nutta et al., 2012). Studies indicate that the majority of PSTs are not receiving adequate training on how to assist students who struggle with language development, which hinders ELs' performance in all content areas (Washburn & Mulcahy, 2014), unlike linguistically aware teachers who consistently gauge and understand their students' language difficulties (Hadjioannou & Hutchinson, 2010). The results of this study indicate that developing PSTs' knowledge about ELs and language development techniques contributes to their success in closing the achievement gap.

Studies show that TPPs must be proactive to meet the needs of ELs in the K–12 system through the inclusion of targeted coursework highlighting EL teaching skills for PSTs (Nutta et al., 2015) by infusing language, grammar, and metalinguistic components in teacher preparation curricula (Svalberg, 2015), at the least. The One-Plus model includes a second language acquisition course designed to improve PSTs' knowledge of language development, seemingly making a difference in how the PSTs address ELs' below average performance in content area assessments.

Program of Study Makes No Difference

The PSTs in this study came from different programs, which required varying level of focus on EL teaching skills. The more infused courses PSTs take, the more they should be qualified to narrow the achievement gap. Yet, the data did not demonstrate a difference in programs. While this was not expected, other research has shown similar findings (Lavery et al., 2018). It is mandatory for all PSTs in the One-Plus model to take two courses addressing EL teaching skills and complete a service-learning project. This result indicates that even at this level of exposure to best practices TPPs can prepare PSTs to have adequate knowledge and skills to work with ELs and are effective in the classroom.

Limitations, Conclusions, and Recommendations

This study contains multiple limitations. First, though the TWS data are strong indicators of teachers' impact on their students, the answers are dependent upon the PSTs filling out the form accurately and truthfully. Second, the data came from a limited subsection of the United States and may not represent the whole country or even the state. Third, different universities use different types of infusion models. The results of one infusion model may not apply to other PST's who attend different teacher programs.

Finally, this is a new area and new program, and it calls for similar studies. The data demonstrate that when PSTs are exposed to EL teaching techniques and inclusive practices, it has a positive impact on EL test scores in their future classrooms. While the One-Plus model is just one of many ESOL infusion models, all TPPs need to investigate the effectiveness of their own EL teaching infusion programs, or lack thereof, for preparing PSTs for the growing diversity of future classrooms. Future researchers should focus their attention on differential achievement gains based on the subject area and One-Plus graduates' personal accounts (qualitative or mixed-method studies) and perhaps a delayed posttest to assess the lingering impact of the infused One-Plus teacher preparation model.

References

Corresponding Author: Nirmal Ghimire
Author Contact Information: ghimire.nirmal@yahoo.com

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment, 8*(3), 231–257. https://doi.org/10.1207/S15326977EA0803_02

Anthony, G., & Walshaw, M. (2009). Characteristics of effective teaching of mathematics: A view from the West. *Journal of Mathematics Education, 2*(2), 147–164.

Babinski, L. M., Amendum, S. J., Knotek, S. E., Sanchez, M., & Malone, P. (2018). Improving young English learners' language and literacy skills through teacher professional development: A randomized controlled trial. *American Educational Research Journal, 55*(1), 117–143. https://doi.org/10.3102%2F0002831217732335

Barrow, L., & Markman-Pithers, L. (2016). Supporting young English learners in the United States. *The Future of Children*, 159–183. https://www.jstor.org/stable/43940586

Beecher, M., & Sweeny, S. M. (2008). Closing the achievement gap with curriculum enrichment and differentiation: One school's story. *Journal of Advanced Academics, 19*(3), 502–530. https://doi.org/10.4219/jaa-2008-815

Benton, J. E., Powell, D., DeLine, M. A., Sautter, A., & Harriet, M. (2012). The teacher works sample: A professional culminating activity that integrates general studies objectives. *The Journal of General Education*, *61*(4), 369–387. https://doi.org/10.5325/jgeneeduc.61.4.0369

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2016). A research synthesis of the association between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research, 87*(2), 425–469. https://doi.org/10.3102%2F0034654316669821

Borman, G. D. Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2005). Success for all: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis, 27*(1), 1–22. https://doi.org/10.3102/01623737027001001

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440. https://doi.org/10.3102/0162373709353129

Carrell, M. R., & Dittrich, J. E. (1978). Equity theory: The recent literature, methodological considerations, and new directions. *Academy of Management Review, 3*(2), 202–210. https://doi.org/10.5465/amr.1978.4294844

Coady, M. R., Harper, C., & de Jong, E. J. (2015). Aiming for equity: Preparing mainstream teachers for inclusion or inclusive classrooms? *TESOL Quarterly, 50*(2), 340–368. https://doi.org/10.1002/tesq.223

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.

Conderman, G., Johnston-Rodriguez, H. P., & Walker, D. (2012). Honoring voicing from beginning special educators for making changes in teacher preparation. *Teacher Education and Special Education, 36*(1), 65–76. https://doi.org/10.1177/0888406412473311

Costa, J., McPhail, G., Smith, J., & Brisk, M. E. (2005). Faculty first: The challenge of infusing the teacher education curriculum with scholarship on English language learners. *Journal of Teacher Education, 56*(2), 104–118. https://doi.org/10.1177/0022487104274119

Darling-Hammond, L. (2008). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57*(2), 120–138. https://doi.org/10.1177/0022487105283796

de Oliveira, L. C., & Athanases, S. Z. (2007). Graduates' reports of advocating for English language learners. *Journal of Teacher Education, 58*(3), 202–215. https://doi.org/10.1177/0022487107299978

Denner, P. R., Norman, A. D., Salzman, S. A., Pankratz, R. S., & Evans, C. S. (2001). Renaissance partnership teacher work sample: Evidence, validity, and quality of student learning assessment. https://www.researchgate.net/profile/Antony_Norman/publication/238111263.pdf

Foster, A., Kohn, L., McGuire, M., Miller, M., & Miller, B. (2010). The power of TWS: Exploring the impact of the Teacher Work Sample (TWS) on teacher candidates' ability to reflect on teaching and learning. *Curriculum and Teaching Dialogue*, 117–136.

Furgione, B., Evans, K., Ghimire, N., Thripp, R., & Russell II, W.B. (2018). What's on your plate?: Correlating subsidized lunch and proficiency on the civics end-of-course assessment. *Educational Practice and Theory, 10*(1), 51–66. https://doi.org/10.7459/ept/40.1.04

Gándara, P., & Maxwell-Jolly, J. (2006). Critical issues in developing the teacher corps for English learners. In K. Tellez & H. C. Waxman (Eds.), *Preparing quality educators for English language learners: Research, Policies, and Practices* (pp. 99–120). Erlbaum.

Gehrke, R. S., & Cocchiarella, M. (2013). Preservice special and general educators' knowledge of inclusion. *Teacher Education and Special Education, 36*(3), 204–216. https://doi.org/10.1177/0888406413495421

Ghimire, N. (2020). Narrowing English learner (EL) achievement gaps: A multilevel analysis of an EL-infused teacher preparation model. *Electronic Theses and Dissertations*, 2020, 216. https://stars.library.ucf.edu/etd2020/216

Goldenberg, C. (2013). Unlocking the research on English learners: What we know—and don't yet know—about effective instruction. *American Educator*, *37*(2), 4.

Good, M. E., Masewicz, S., & Vogel, L. (2010). Latino English language learners: Bridging achievement and cultural gaps between schools and families. *Journal of Latinos and Education, 9*(4), 321–339. https://doi.org/10.1080/15348431.2010.491048

Hadjioannou, X., & Hutchinson, M. C. (2010). Putting the "G" back in English: Preparing pre-service teachers to teach grammar. *English Teaching: Practice and Critique*, *9*(3), 90–105.

Hunter, R. C., & Bartee, R. (2003). The achievement gap: Issues of competition, class, and race. *Education and Urban Society*, *35*(2), 151–160. https://doi.org/10.1177/0013124502239389

Hutchinson, C. J. (2011). Infusing EL content into a foundations course. *The Tapestry Journal, 3*(2), 13–18.

Isaac, L. B., & Quantz, R. A. (2011). Infusing EL content into a sociocultural study in education course. *The Tapestry Journal, 3*(2), 7–12.

Jaynes, W. H. (2015). A meta-analysis on the factors that best reduce the achievement gap. *Education and Urban Society, 47*(5), 523–554. https://doi.org/10.1177/0013124514529155

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher, 39*(6), 484–486. https://doi.org/10.3102/0013189X10378400

Lancour, M., & Tissington, L. D. (2011). The effects of poverty on academic achievement. *Educational Research and Reviews*, *6*(7), 522–527. https://doi.org/10.5897/ERR.9000349

Lavery, M. R., Nutta, J., & Youngblood, A. (2018). Analyzing student learning gains to evaluate differentiated teacher preparation for fostering English learners' achievement in linguistically diverse classrooms. *Journal of Teacher Education*, *70*(4), 372–387. https://doi.org/10.1177/0022487117751400

Lavy, I., & Maschiach-Eizenberg, M. (2017). The interplay between spoken language and informal definition of statistical concepts. *Journal of Statics Education, 17*(1). https://doi.org/10.1080/10691898.2009.11889502

McClain, K., & Cobb, P. (2004). The critical role of institutional context in teacher development. *International Group for the Psychology of Mathematics Education, 2004.* https://eric.ed.gov/?id=ED489583

McCoach, D. B. (2010). Hierarchical linear modeling. In G. R. Hancock & R. O. Muller (Eds.), *The reviewer's guide to quantitative methods in the social sciences*, 123–140. New York, NY: Taylor & Francis. https://doi.org/10.4324/9781315755649

McHatton, P. A., & Parker, A. (2013). Purposeful preparation: Longitudinally exploring inclusion attitudes of general and special education pre-service teachers. *Teacher Education and Special Education, 36*(3), 186–203. https://doi.org/10.1177/0888406413491611

Meskill, C. (2005). Infusing English language learner issues throughout professional educator curricula: The training all teachers' project. *Teachers College Record, 107*(4), 739–756.

Niday, D. (2012). Infusing EL content and instruction into English education courses. *The Tapestry Journal, 3*(2), 19–25.

Nutta, J. W., Mokhtari, K., & Strebel, C. (2012). *Preparing every teacher to reach English learners*. Harvard Education Press.

Nutta, J. W., Strebel, C., Mokhtari, K., Mihai, F. M., & Creveocoeur-Bryant, E. (2015). *Educating English learners: What every classroom teacher needs to know* (2nd ed.). Harvard Education Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Smith, P. C. (2011). Teaching inclusivity: Preservice teachers' perceptions of their knowledge, skills, and attitude toward working with English language learners in mainstream classrooms. *The Tapestry Journal, 3*(1), 1–21.

Svalberg, A. M. (2015). Understanding the complex processes in developing student teachers' knowledge about grammar. *Modern Language Journal*, *99*(3), 529–545. https://doi.org/10.1111/modl.12241

U.S. Census Bureau. (2020). *Quick facts- Hispanic or Latino, percent.* Retrieved from https://www.census.gov/quickfacts/fact/table/US/RHI725221

van Laarhoven, T. R., Munk, D. D., Lynch, K., Bosma, J., & Rouse, J. (2007). A model for preparing special and general education preservice teachers for inclusive education. *Journal of Teacher Education, 58*(5), 440–455. https://doi.org/10.1177/0022487107306803

Washburn, E. K., & Mulcahy, C. A. (2014). Expanding preservice teachers' knowledge of the English language: Recommendations for teacher educators. *Reading & Writing Quarterly*, *30*(4), 328–347. https://doi.org/10.1080/10573569.2013.819180

Welsh, R. O. (2018). Student mobility, segregation, and achievement gaps: Evidence from Clark County, Nevada. *Urban Education, 53*(1), 55–85. https://doi.org/10.1177/0042085916660349

Wheeler, D. L., & Govoni, J. M. (2014). An ESOL curricular model: Infusing ESOL standards in teacher education. *The Tapestry Journal, 6*(2), 10–35.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461. https://doi.org/10.1037/0033-2909.91.3.461

Yoder, P. J., Kibler, A., & van Hover, S. (2016). Instruction for English language learners in the social studies classroom: A meta-synthesis. *Social Studies Research and Practice, 11*(1), 20–39.

Zinth, J. D. (2013). Career/technical education: Not your father's vocational education. *The Progress of Education Reform, 14*(1), 1–7.