2020

Narrowing English Leaner (EL) Achievement Gaps: A Multilevel Analysis of an EL-infused Teacher Preparation Model

Nirmal Ghimire
University of Central Florida

NARROWING ENGLISH LEARNER (EL) ACHIEVEMENT GAPS: A MULTILEVEL ANALYSIS OF AN EL-INFUSED TEACHER PREPARATION MODEL

by

NIRMAL GHIMIRE
B. A. Tribhuvan University, 2007
M.A. Tribhuvan University, 2010

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the School of Teacher Education
in the College of Community Innovation and Education
at the University of Central Florida
Orlando, Florida

Summer Term
2020

Major Professor: Joyce W. Nutta

ABSTRACT

This non-experimental correlational study assessed the effectiveness of a model of teacher preparation that infuses a focus on teaching English learners, the One-Plus model, by examining pre-service teachers' (PST) effectiveness in narrowing English learner achievement gaps during their final student teaching experience. The study spanned five semesters of internship data, exploring how each semester's PST effectiveness changed over time. This study utilized teacher work sample data that interns collected and submitted to the institutional effectiveness division of the college, with an *n* of 20,809 K-12 students who attended the 768 One-Plus PSTs' classes during their semester-long internship. The results showed that there remained a statistically significant achievement gap between student groups based on their sociodemographic characteristics, and the biggest gap was between EL and non-EL students. Students had statistically significantly higher posttest scores compared to the pretest scores, and the rate of change in test scores was much steeper in historically low-achieving students than their counterparts. There was an approximately 50% decrease in EL's achievement gap in the posttest model compared to the pretest. Likewise, the gaps between low-SES and high-SES students, students with exceptionalities and students without exceptionalities, Black and White Students, and Hispanic and White students were reduced by approximately 40%, 38%, 48%, and 26%, respectively. Finally, there was a statistically significant linear growth in students' posttest scores over a period of five semesters.

Key Words: teacher education, infusion model, One-Plus model, hierarchical linear model (HLM), English learners (ELs), preservice teachers, three-level hlm, preservice teacher education

I want to dedicate this dissertation to my dad Lekh Nath Ghimire and mom Kamala Maya Ghimire who could not make it to my graduation because of the ongoing pandemic. It has been ten long years, and we all were very excited that we would get to meet on this occasion. You made this day possible. Thank you!

ACKNOWLEDGEMENTS

I feel incredibly honored to have reached this point in life. It was a dream of a lifetime for me to work with prominent authors, researchers, education enthusiasts, and great teachers/mentors with unparalleled expertise and skills. The University of Central Florida made every single dream come true and many more. My sincere and earnest thanks go to Dr. Joyce W Nutta for her unwavering support, guidance, and encouragement. Thanks for dedicating your whole life to designing something so beautiful in the form of the One-Plus Teacher Education program, which for sure will fulfill your dream of equity education and equitable learning gains among all students!

I am immensely thankful to Dr. Florin M Mihai for his temperament, attitude, and leadership for 4-years as my mentor and supervisor for TSL 4240. I got to learn much more from you and the teaching opportunities than what I ended up giving to my students. You are one of the best professors I ever met. I hope you will be able to touch the lives of many graduate students in the coming days and years. Dr. Donita Grissom, you are an epitome of compassion; without it, a teacher becomes a non-teacher. Thanks for your expertise and help throughout this process. I enjoyed taking your classes and working with you in the MELTS project. Dr. Shiva Jahani, you were the only go-to person at UCF. Your statistical knowledge is exceptional, and you allow it to be available for every graduate student who comes your way. I want to acknowledge Dr. Orin Smith and his help in getting access to data. Without your help, it would have been impossible. I appreciated your attitude and support. Thanks.

To my dad and mom, I cannot express enough how lucky I am to have had you as my parents. Your unconditional love and enthusiasm towards my dream encourage me always to accept challenges. Thanks for everything you gave to my family and me.

To my wife, Sushila, I cannot thank you enough for all your dedication and support throughout my life and especially through this journey. You have taught me when to pause and when to proceed and how to achieve the monumental dreams that, at times, look beyond our reach. This moment would not be possible without your company, sacrifice, and perseverance. Thanks for using your shoulder to lift the whole burden of our precious daughter. To my daughter, Aanishma, I do not know where to begin. It is a joy to have you around. I love your silly antics! You spark a glimpse of inspiration and purpose during the dark moments. You encourage me to be a better person day in, day out. Thanks for being who you are.

To my siblings, Suvash, Samjhana, and Sushil, I appreciated your support and encouragement, thus far. You have been a critical part of this journey. This walk could have been much harder had you not provided the financial and psychological support. To all my relatives and my in-laws, Ambika Regmi and Shankar Regmi, thanks for your constant support and encouragement. I acknowledge your portion in this achievement, and I appreciate all for your time and prayers on our behalf.

My colleagues at the University of Central Florida, Lauren, Xiong (Charlene), Antonio, and Shizhong were the best among all. Thanks for sharing your stories and listening to mine. You inspired me to overcome hardship and stay focused. Thanks for all your support. Finally, thanks to all UCF professors, students who attended my classes and taught me some valuable

lessons of my life, staff in the College of Arts and Humanities, and everyone who touched my life through this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criteria
ANOVA	Analysis of Variance
BICS	Basic Interpersonal Communication Skills
CCSD	Clark County School District
EL	English Learners
ELD	English Language Development
ELL	English Language Learners
ESL	English as a Second Language
FIML	Full Information Maximum Likelihood
FRPL	Free or Reduced-Price Lunch
HLM	Hierarchical Linear Modeling
HSPE	High School Proficiency Exam
ICC	Intra-class Correlation
IRB	Institutional Review Board
K	Kindergarten
L1	First Language
L2	Second Language
LEP	Limited English Proficient
MANCOVA	Multivariate Analysis of Covariance
MET	Measuring Teachers’ Effectiveness

NAEP	National Assessment of Educational Progress
NCES	National Center for Education Statistics
NCLB	No Child Left Behind
OLS	Ordinary Least Squares
PreK	Pre-Kindergarten
PST	Preservice Teacher
SASS	Schools and Staffing Surveys
SD	Standard Deviation
SES	Socio-economic Status
SLA	Second Language Acquisition
SPSS	Statistical Package for the Social Sciences
STEM	Science, Technology, Engineering, and Mathematics
TESOL	Teaching English to Speakers of Other Language
TPP	Teacher Preparation Program
TWS	Teacher Work Sample
UCF	University of Central Florida
U. S.	United States
VAM	Value-Added Model
WIDA	World-Class Instructional Design and Assessment

CHAPTER ONE:
INTRODUCTION

Statement of the Problem

The achievement gap can be defined as inequality in educational outcomes among student groups (Anderson, Medrich, & Fowler, 2007). Such gaps come into being when a group of students categorized by race/ethnicity, gender, socioeconomic status, type of school, etc. performs statistically significantly better than another group and the difference in mean scores for the two groups exceeds the margin of errors (National Center for Education Statistics, 2015). Among all the student groups, English learners (ELs) are a historically lowest-performing subgroup, and they are "often among the most vulnerable to academic failure and therefore they are the most in need of academic supports" (Sanders, Durbin, Anderson, Fogarty, Giraldo-Garcia, & Voight, 2018, p. 658).

The K-12 education system has always been marred by a statistically significant achievement gap between EL and non-EL students (Banks, 2012: Welch-Ross, 2010). Until the last decade, many ELs participated in English for speakers of other languages (ESOL) classes taught by professionals with specialized language and pedagogical skills (e.g., ESOL specialists), with the assumption that this "pull-out" time from their general classroom instruction would address all their learning needs. Although this specialized, separate form of support provided many benefits, it could have the unintended effect of removing generalist classroom teachers' sense of responsibility for their English learners' performance if they perceived their ELs as the purview of ESOL specialists. The system of providing dedicated English for Speakers of Other Languages (ESOL) teachers in schools had the potential to hinder generalist teachers from acquiring foundational knowledge about EL issues, let alone successfully running inclusive classrooms (McHatton & Parker, 2013; Coady, Harper, & de Jong, 2015). Thus, many in-service teachers, and especially content area teachers, did not feel adequately prepared to teach English learners (ELs) and admitted that they lacked the knowledge and tools to help ELs (Lavery, Nutta, & Youngblood, 2019; Zinth, 2013).

Based on this assertion, the repetitive pattern of significant differences in learning gains despite the untiring efforts by educators, education restructuring initiatives, state and local initiatives, the appropriations of federal grants, and an enormous body of educational theory, research, and annotation (Troff, 2014), academic achievement gaps persist between ELs and non-EL students. At times, the contention that education is an equalizer in the face of structural disparities appears to be falling apart along students' racial, ethnic, linguistic, and socioeconomic status (Hung, Smith, Voss, Franklin, Gu, & Bounsanga, 2019). In addition, further research provided key insights into the fact that teachers who have little to no experience with ELs tend to believe that ELs are unable to benefit from higher-order challenges, and they deserve remedial curricula, which further lowers the degree of challenges and expands the gaps (Troff, 2014).

Generalist teachers who know something about language are believed to be able to effectively help ELs, even though there is not a clear demarcation of language components that the mainstream teachers need to learn, nor how teacher preparation programs (TPPs) can best prepare teachers to develop this knowledge. However, it is not surprising that essential knowledge regarding language and linguistics is not considered to be of prime importance in their professional practice by most mainstream teachers (Bunch, 2013). The quality of instruction is believed to be more important than the language of instruction, hence, ELs are not able to get quality teachers as most of the mainstream teachers do not have proper training in ESOL teaching methods and/or cultural competency (Wixon, 2015). Thus, the proposition of new academic standards in states with high numbers of EL students attempted to make sure that teacher preparation programs and states' professional development programs properly prepare educators to become capable of supporting ELs through the language development and academic support to help them succeed. States are recommended to base their licensure advances and tenure decisions based on teacher effectiveness, allow fair and expedient processes to terminate the ineffective teachers, consider classroom performance in layoff decisions, and direct leadership opportunities toward the retention of effective teachers to boost the quality of teacher workforce (Mead & Eckes, 2018).

Infused ESOL Model of Teacher Preparation

The findings of various studies on the achievement gap and equity education have put teacher education in the center of the discussion. Some educators assert that focus on pre-service teacher education can effectively help narrow the achievement gap between English learners and their counterparts in mainstream American classrooms (Nutta, Mokhtari, & Strebel, 2012; Maulana, Helms-Lorenz, & van de Grift, 2015; Goddard, Sweetland, & Hoy, 2000). Thus, the pursuit of effective teaching and quality learning within the era of high teacher accountability has brought us to a recent development known as an infused teacher preparation model, which maintains that integration of a focus on ELs throughout the curriculum is necessary to prepare teachers and accelerate the amount and quality of ELs' second language development, thereby narrowing the EL, non-EL gaps (Hunter & Bartee, 2003).

Proponents of infused ESOL TPPs assert that there is a need for systematic and broad curricular development that contributes to all generalist pre-service teachers' (PSTs) preparation to teach and assess ELs (Goldenberg, 2008). They maintain that PSTs would be able to apply their acquired knowledge upon successful implementation of the ESOL infused TPP that provides a clear elicitation of ESOL standards in teacher education curricula (Wheeler & Govoni, 2014).

One infused ESOL teacher preparation approach, the One-Plus model, is among a number of infused ESOL initiatives and a recent development in the teacher education arena. This flexible teacher education model offers four different outcomes, (a) Basic Coverage, (b) EL-qualified for academic subjects, (c) EL-qualified for language arts, and (d) English as a Second Language (ESL) endorsement or certification, which require variable degrees of input elements (Nutta et al., 2012). For example, in programs that prepare PSTs to become EL-qualified for teaching academic subjects, most of the teacher education courses have embedded content and assignments that focus on ELs compared to only a few EL-embedded courses for the PSTs in the basic coverage option. Another contrast is the number of EL-specific courses that are part of the ESOL-infused program. Basic coverage, for example, does not include any EL-specific coursework and relies solely on what can be embedded into existing courses in the teacher preparation curriculum. For programs leading to the EL-qualified for language arts outcome, the majority of the teacher education courses for language arts PSTs, are broadly EL embedded, and additional EL-specific courses are incorporated, while PSTs who pursue the ESL endorsement or certification progress across thoroughly EL embedded courses take additional EL-specific coursework (Nutta et al., 2012, p. 48).

The One-Plus TPP maintains that a "combination of program components, including EL-embedded and EL-specific courses with associated field/clinical experiences and systematic formative and summative candidate evaluation improve the quality and impact of the infusion effort" (Nutta et al., 2012, p. 52). A limited body of research has measured the effectiveness of infusion and has reported higher learning gains and improved teacher preparation (Coady et al., 2015; Lavery et al., 2019). In lieu of small-scale studies that attempted to identify the impact of a single component of the infused model, the infusion process, or qualitative research that looked into the lived experiences of few educators and/or PSTs, a large-scale data-driven study is warranted to gauge the outcome of these PSTs. Moreover, not a single study explores the effectiveness trend for an extended period of time using multiple students and teacher-related variables.

Evolving Interdisciplinary Theory of Teacher Education

Education is widely linked to the economy; thus, the production of well-qualified teacher force—prepared to provide a world-class standard for all students, including ELs, and to narrow the achievement gap and shift with educational reforms—is key to sustaining nations' posture in the global economy. Evidently, the issues like how teachers are prepared, how it relates to public policy, and how to maintain social justice in classrooms have been in the crux of educational discussion for decades (Decker, Rice, & Moore, 1997). In addition, the rapid change in K-12 population demographics in the United States calls for professional development requirements for in-service teachers and new TPPs for preservice teachers. New types of TPPs have begun to appear focusing on democratic education, multicultural education, equity education, or teachers' beliefs and identity, which compelled traditional teacher education programs to modify their existing curricula with newfound aspects (e.g., service-learning, etc.) and embed community partnership and technology within their models (Cochran-Smith, 2001). All these models strive to foster PSTs' knowledge of inclusivity and cultural responsiveness.

The traditional notion that required ELs to learn English in ESOL classrooms from highly specialized educators and get exposure to the English language in content area classrooms did not yield the desired outcome because of the ineffectiveness, and innately confusing nature of educational policies (Raubaugh, 2019); or lack of enough manpower to work with growing number of ELs (Darling-Hammond, 2006b), and lack of support and nurturing educational environment for ELs in content area classrooms (Nutta et al., 2012). Thus, the new trend in teacher education propelled the idea of preparing all teachers (generalist teachers) to help ELs develop both content-area knowledge and language knowledge simultaneously, regardless of their subject areas (Nutta, Strebel, Mihai, Crevecoeur-Bryant, & Mokhtari, 2018). The primary findings of these TPPs provide encouraging results in that graduates feel more prepared and confident in working with all students, including ELs, which strengthened the idea that highly qualified and professional teaching force can be prepared through teacher education (Darling-Hammond, 2006).

To prepare preservice teachers to meet daunting standards, a new theory of teacher education to lay out the foundation for a new teacher preparation program was warranted—a theory that can identify the components of traditional teacher preparation models that are working, and building on them to promote language acquisition and higher achievement for English learners. As teacher preparation programs cover a wide range of content areas, the required theory had to be interdisciplinary in nature to critically evaluate disciplinary insights and create common ground among them to assure more comprehensive understandings (Repko & Szostak, 2017). Standing on the foundation of the Interactionist perspective of Second Language Acquisition (SLA), the Infused ESOL One-Plus model aims toward filling the gap in teacher education philosophy.

As in the most critical interdisciplinary areas, the One-Plus model serves the purpose of dismantling the boundaries between existing teacher preparation practices among various pedagogical content areas, preparing teacher candidates to treat cultural issues relationally, and encouraging to advocate and practice inclusion of a marginalized student body, e.g., ELs. Built on the foundation of empirical research, literature, and classroom practices, the One-Plus model assumes that the underlying learning theories between teacher education of various subjects (e.g., math education and English language arts) present evidence of inherent interconnectedness between them. Common differences include the focus, emphases, and assumptions, which may seem to be lying somewhere on a continuum (Nutta et al., 2012). The issues are thus, complex, yet, the goal of successful infusion is attainable if the process is followed as outlined in its blueprint (Repko & Szostak, 2017).

To accomplish the daunting task of infusion of ESOL components in content area courses, the One-Plus model offers measures that include 'reflective portfolios, benchmark reviews of candidates' developing knowledge and skills at key points in the program of studies, cumulative summative assessments, and capstone experiences with ELs, and among others' (Nutta et al., 2012, p. 21). The goal is achievable only if the stakeholders (e.g., faculty mentors, ESL faculty, infusion team, the host institutions, etc.) work together in good faith and stick through the long repetitious processes in all five key phases, i.e., the assessment, design, development, implementation, and evaluation phases (Moallem, 1998). Departing from the traditional content area-based teacher education programs towards the One-Plus model offers what its proponents call ‘unique opportunities to cultivate cohesion in candidates’ perception of ELs and how to teach and assess them in comprehensible and equitable ways’ (Nutta et al., 2012, p. 21).

Purpose of the Study

The purpose of this study is to assess One-Plus PSTs’ preparedness to bridge the achievement gap between EL and non-EL K-12 students in the United States school system. As a comprehensive TPP system, the One-Plus model offers a wide range of knowledge bases on pedagogy, assessment, curriculum, and technology for its PSTs. In addition, it provides multiple hands-on-training opportunities through activities like service-learning, micro-teaching, and teaching to digital avatars in simulation and clinical settings for its PSTs. Such activities are geared toward gaining experience with ELs, appreciating and understanding EL issues, and reflecting on possible solutions. Based on these assertions and a few preliminary findings, the One Plus PSTs were predicted to be able to narrow the EL achievement gap. In this context, this study advances our understanding of this field a step further by delineating if the One-Plus TPP is successfully producing capable teachers who effectively narrow the EL achievement gap. Thus, this study was designed to offer a useful insight into the PSTs’ individual and collective effectiveness in limiting and/or eliminating the achievement gap, as an outcome of the TPP they attended.

Research Questions

The following research questions guided this study:

1. What is the rate of change of One-Plus PSTs' effectiveness in reducing the EL achievement gap over time?

    1.1. What proportion of the variance in students' academic achievement is associated with students' EL-status, FRPL-status, ethnicity, disability status, and gender?

    1.2. What proportion of the variance in students' academic achievement is associated with PSTs' major, grade-level of teaching, class size and the subject they teach?

A quantitative correlational design was used to analyze the pretest and posttest unit assessment scores of the K-12 students in the classrooms of the One-Plus PSTs for the semesters of Fall 2016, Spring 2017, Fall 2017, Spring 2018, and Fall 2018. The data are a part of Teacher Work Sample (TWS) required during PSTs' final internship using an Excel-based spreadsheet known as GraphMaker™ version 5.1.2 (Appendix B, Part 2).

PSTs who complete a full-time semester-long student teaching internship made up the key population of the inquiry. This study uses the value-added model of teacher effectiveness, i.e., the statistical system that accounts for the differences in students' test scores before and after instruction (Kupermintz, 2003), thus all the students taught by the One-Plus PSTs during their internship regardless of their grade-level are other important population in this study.

Based on the type of data and the inquiry sought forth by the research questions, three separate procedures are necessary. First, a three-level hierarchical design is considered most appropriate to assess the impact of student and PST level variables individually (main effect) and how they change by semester. Second, a Multivariate Analysis of Covariance (MANCOVA) is identified to measure the change in English learners' pretest and posttest scores as a function of their socioeconomic status, gender, disability status, ethnicity including the PST level variables (e.g., major, subject, the grades the PSTs taught, and the class size). Finally, a two-level longitudinal hierarchical linear modeling (HLM) is selected to study the change in students' pre- and posttest scores over five semesters.

Significance of the Study

This study could be significant should the findings demonstrate that the PSTs who completed the infused ESOL teacher preparation program prompted the higher learning gains in all students, and more so, in low-achieving students including the ELs, and that the gain remains either consistent or keeps growing over time. This study will make several contributions to educational policymakers, state or federal departments of education, university teacher preparation programs, and language teaching programs.

First, this study serves as a foundation for further exploratory analyses of Infused ESOL (the Florida Model) and Infused ESOL One-Plus teacher preparation models and the effectiveness of the teacher preparation model to enable comparisons among various available models and make informed decisions about adapting, modifying, or continuing existing TPPs in university, state, and federal levels. In addition, the state and federal education policymakers are better served by most recent longitudinal research findings to help them embed best practices into the educational policies to guarantee effective TPPs and relevant components elsewhere with a higher degree of fidelity.

Likewise, universities and teacher preparation schools obtain access to new information about the application of the Infused ESOL One-Plus model, be able to compare the results with components of their existing programs, and help them decide about possible adaption or adaptation. Educators, teacher educators, or universities are encouraged to conduct a thorough data-driven evaluation of existing TPPs in terms of PSTs' current status, the average length of teaching, and whether they can meet their students' everyday classroom needs.

Furthermore, preservice teacher candidates benefit most from this study of the implementation of ESOL Infusion One plus model in teacher education programs because the results in this study help enhance their opportunities to develop the appropriate knowledge base to understand and reach all students regardless of their demographic, socioeconomic, and linguistic differences, and be successful in their heterogeneous classrooms. Furthermore, they have the potential to learn the critical pedagogical skills relevant to ELs, finding appropriate textual and technological resources, and using them properly to narrow the learning gaps.

The findings of this study may be significant for teacher trainers and professional development agencies to design new professional development courses or reform the existing courses to enhance in-service teachers' capacity to work with ELs. If so, the current achievement gap may be decreased, thereby fulfilling the promise of equitable education.

Definition of Major Terms

Achievement Gaps. Achievement gap originally referred to inequality in educational outcomes between White and Hispanic students (Anderson et al., 2007), which further extended to the concept of unequal academic achievement between the students of different ethnic, linguistic, and racial backgrounds (Banks, 2012; National Center for Education Statistics, 2015). Furthermore, the achievement gap exists if a group of students performs better than other groups, and the disparity in mean scores for the two groups is statistically significant (Anderson et al., 2007).

English Language Proficiency (ELP). It refers to English learners' ability to use the English language to make and transmit meaning in spoken and written contexts (University of Southern Queensland, 2016) assessed by the tests that measure ELs' general language abilities. The tests are the key assessments to identify if a student requires ESOL placement, reclassification, or departure from this placement (Linquanti & Cook, 2013).

English Learner (EL). The United States Federal law defines English learner as someone whose native language is other than English, who is aged between 3-21 and enrolled in an elementary or secondary school and possesses sufficient difficulties in speaking, reading, writing, or understanding the English language to deny the individual, (a) the ability to meet the State's proficient level of achievement on State assessments, (b) the ability to successfully achieve in classrooms where the language of instruction is English, or (c) the opportunity to participate fully in society (Education Commission of the States, 2014). In this study, this term refers to a US K-12 active learner of the English language who attended the One-Plus PSTs during their internship in a mainstream class. This term may have been used interchangeably with English language learners (ELL) or EL students in this study while quoting studies that use these terms (National Council of Teachers of English, 2008).

English for the Speakers of Other Languages (ESOL). In this study, this term is used to refer to the program of study that provides focused literacy blended instruction intending to make

ELs fully proficient in English and meet the challenging state and federal standards (Santiago, Martinez, & Peter, 2016). For this study, ESOL refers to the teacher preparation program that embeds EL instructional strategies that require PSTs to receive some form of tailored instruction with a specially designed curriculum targeted to ELs (Nutta et al., 2012).

Preservice Teacher (PST) and Teacher Preparation Programs' (TPPs) Effectiveness. This term is defined mainly as students' achievement on standardized tests, along with students' observations and surveys (Cochran-Smith, Keefe, Chang, & Carney, 2018). In this study, PSTs' effectiveness is the average achievement of all their students on pretest and posttest, while the TPPs' effectiveness is the average achievement of all PSTs on their students' pretest and posttest scores.

First Language (L1). This term refers to a language that a child acquires in different stages through natural language processing using biologically endowed mechanisms like Universal Grammar (VanPatten & Benati, 2015). It can be any language that a child is exposed to, during early linguistic development and continues to be exposed to through parents, community, or any other media.

Free and Reduced-Price Lunch (FRPL). A federally assisted School Lunch Program that offers K-12 students a free lunch who come from the family with income at or below 130 percent of the Federal poverty level, and reduced-price lunch for those who belong to the household incomes between 130-185 percent of the federal poverty level (United States Department of Agriculture, Food and Nutrition Services, 2017). Many researchers use this indicator as a proxy for students' socioeconomic status (Furgione, Evans, Ghimire, Thripp, Russell, 2018). In this study, the FRPL-eligible students refer to low-SES students and vice-versa.

Infusion Model: This term refers to a specially designed teacher preparation program that aims to prepare content area teachers to instruct and assess ELs in the mainstream classroom through the supplement 'of EL content in an interconnected, cohesive, and interdisciplinary manner' (Nutta et al., 2012, p. 26). In general, Infusion model is a teacher education initiative in which the generalist teacher candidates undertake an expanded program designed to meet the guidelines and standards to work with all students including ELs (van Laarhoven et al., 2007).

One-Plus Model: It is a unique Infusion model of teacher preparation program that is an extension of the general infusion model, which is flexible to meet varying requirements of teaching and non-teaching professionals. This model aims to improve the quality and impact of PSTs through a series of EL-embedded and EL-specific courses with associated field/clinical experiences and systematic formative and summative evaluation (Nutta et al., 2012). The stackable nature of this model allows teacher candidates to pursue differential level of rigor and depth based on their requirement, resources, and academic degree areas. For example, the 1+ EL-embedded courses focus on the learners and learning context, while the 2+ embedded courses focus on planning and implementing curriculum, instruction, and assessment in content areas. Finally, the 3+ embedded courses geared towards planning and implementing curriculum, instruction, and assessment in language arts and literacy (Nutta et al., 2012).

Preservice Teacher (PST). A participant in this study who attended the infused ESOL One-Plus model of the teacher preparation program and completed student teaching in a mainstream class that included ELs. This person is responsible for completing a teacher work sample as part of the requirements for an advanced or continuing license. Sometimes the person is also called *a student teacher* or *a teacher candidate*.

Second Language (L2). The term second language (L2) is defined as any language learned after learning the first language (Gass & Selinker, 2008). In practice, this term also refers to the language somebody is learning, i.e., the target language (TL), even if it is their third or fourth language (or more) (Ellis, 2015).

Student. Cambridge dictionary defines a *student* as 'a person who is studying at a school, college, or university.' In this study, this term exclusively refers to a K-12 learner taught by a Preservice teacher during their internship.

Teacher Preparation Program (TPP). This term refers to a university-based program that is dedicated to producing future teachers through a set of courses and experiences. The current trend in TPPs aims to train teachers as classroom researchers and expert collaborators who can help a diverse set of students and their infinitely diverse learning ways (Darling-Hammond, 2006b).

Teacher Work Sample (TWS). The product was initially developed at Western Oregon University to document preservice teachers' level of competency to be eligible for licensure. It has two portions, (a) qualitative description of the learning context and the instructional unit including learning goals and pre- & post-tests, and (b) GraphMaker™ (Version 5.1.2), a generic Microsoft Excel-based Software designed by Lavery (2012) to record students' demographic and test information.

Assumptions, Limitations, and Delimitations

Delimitations

The participants in this study came from the tracks of teacher preparation programs that included Elementary Education, Math Education, Science Education, English Language Arts Education, Social Studies Education, Art Education, and Foreign Language Development Education. All participants completed their student teaching practicum. Any participant who did not complete the internship was not included in the study. Limitations are discussed at length in Chapter 5.

Assumptions

a. The PSTs were educated that the same tests could cause testing effects, and they followed the test protocol, i.e., the tests "measure the same learning outcomes at the same level of difficulty using similar assessment items and/or methods and produce scores that can be compared to one another to show growth" outlined in TWS manual (Test protocol, Graphmaker™, Version 5.1.2.; Lavery, 2012).
b. To ensure that the changes/gains in the posttest scores be reasonably attributed to student learning and not to the memory of the pre-assessment, the PSTs followed the instruction to make relatively subtle changes to their assessment (Test protocol, Graphmaker™, Version 5.1.2.; Lavery, 2012).
c. The PSTs were honest in their responses and completed the GraphMaker™ (version 5.1.2) wholly and truthfully.
d. The PSTs participated in TWS training and followed the recommendations before they filled out the GraphMaker™ form.
e. The PSTs took sole responsibility for designing learning goals and objectives, creating assessments, and grading them.

Organization of the Dissertation

This dissertation is organized into five chapters. Chapter 1 presents the introduction of the study, the background of the problem, the problem statement, the research question and sub-questions, significance of the study, including a brief overview of the design of the study, and delimitations and assumptions.

The next chapter in this study discusses the issues about the achievement gap, teacher effectiveness, reforms in teacher education to address the achievement gaps, an Infused ESOL model of teacher preparation programs, and its preliminary findings to lay out the case for the importance of this study. The key theme of the infused ESOL One-Plus model is to address the achievement gap head-on and provide a viable option of teacher preparation that is flexible, adaptable, and yet, effective in delivering required knowledge base and tools to reach out to all students. As such, this literature review provides additional insight into the Infused ESOL One-Plus model of teacher education. The analytic focus on the achievement gap offers another insight. This study analyzes the status of achievement gap based on students' EL and non-EL status, SES, students' grade level, class size, disability status, and subject matter.

Chapter 3 delineates the proposed methodology of this study in detail. It further explains the research design and the usefulness of the teacher work sample (TWS) data in evaluating pre-service teachers' performance effectiveness. It elucidates the site, data collection procedure, sample size, variables of interest and their coding, proposed model in each level, and analytical strategy.

Chapter 4 presents the model building process and how the final model was identified and why it was the best fitting model. In addition, it presents the findings vis-à-vis the research question and sub-questions. It further elaborates on the technique and procedure taken to ensure the credibility and generalizability of the findings. Chapter 5 explores the further meanings of the results and their practical implications. A complete set of limitations and recommendations for future research make up the final components in this chapter.

Conclusion

The achievement gap between student groups has long been present in the United States K-12 education system. Linguistic minority students, students with exceptionalities, students from lower socioeconomic status, and ethnic minority students score significantly lower than their counterparts in high-stakes tests that determine their promotion to upper classes and funding for the schools they attend. This issue has been looked at from different perspectives, and there have been multiple attempts to narrow, let alone eliminate such gaps. However, the problem persists. The contemporary research in this direction points towards teacher preparation programs as the primary means to tackle the achievement gap (Darling-Hammond, 2000). A broad consensus is that the production of a small subset of dedicated ESOL force is not a solution, instead, the whole teacher task force should be equipped with the required knowledge base, training, and expertise to help all students efficiently in their classroom (Nutta et al., 2012; Lavery et al., 2019). An Infused ESOL One-Plus model can be instrumental in this scenario. All in all, this study sheds light on the One-Plus PSTs' effectiveness in narrowing academic achievement gaps among students and how the trend remained over five semesters.

CHAPTER TWO:
LITERATURE REVIEW

This chapter will review and analyze the research and scholarship on preservice teacher (PST) education programs designed to narrow the achievement gap between EL and non-EL K-12 students in public schools in the United States. Although studies in teachers' impact on students' academic gain have examined the effectiveness of the use of technology, teaching practices, and curriculum reform on PST education, these studies have not studied teacher effectiveness in consideration with the overall teacher education program and its components. As such, this literature review provides additional insight into the ESOL Infusion One Plus model of teacher education. The analytic focus on the achievement gap offers another insight. This study analyzes the status of achievement gap based on students' EL and non-EL status, SES, students' grade level, class size, disability status, and subject matter.

Though numerous studies in teachers' impact on students' learning gains have established its sizeable effect, little analytic attention has been paid how likely the teacher-level variables are to moderate such growth, to what extent, and how likely do they interact with various student-level variables. This study addresses this issue by demonstrating an exhaustive review of relevant citations.

An initial literature search was conducted to develop benchmarks for identifying studies to incorporate in this review. A search of all relevant literature published in English language journals in four foremost areas used the following phrases, (a) *achievement gap between English learners and non-English learners*, (b) *teachers' impact on students' learning gains*, (c) *teacher education and achievement gap*, and (d) *ESOL Infusion One Plus model*. Also, A list of terms was compiled (*achievement gap*, *English language learners*, *non-ELLS*, *value-added measure*, *preservice teacher*, *teacher training*, *PST shifts*, *Infusion*, *multicultural education*) and a computerized literature search on ERIC, and PsycINFO database from 2010 through 2018 was conducted. The studies that were not related to the United States were excluded from the study, and a higher importance was given to peer-reviewed journal articles, full text, thesis/dissertation, conference proceedings, and reports. Once this study identified the key journal, i.e., the Journal of Teacher Education, a thorough search of the related article was conducted in its homepage. After gathering suitable materials from the education databases, this study searched through the reference lists and used Google Scholar to find similar studies that may have been overlooked in the initial search. In addition to journal articles, the ProQuest Dissertations and Theses database were searched for related dissertations or theses. Searching in the title for the term (*ESOL Infusion Model) and (Teacher Education)* resulted in 512 dissertations while selecting full text within the last two years yielded resulted in 38. On reading the abstracts, there were only eight dissertations and theses connecting culturally responsive education (CRE) to student success that had not been turned into journal articles found in previous searches. Last, library database for books published in the last ten years that included chapters relating to such topics and skimmed summaries to find relevant examples. After narrowing down the seminal works in the areas of interest, additional articles were located through the reference section of those works (Hadwin, & Winne, 2012).

No research synthesis is exhaustive (Boote & Beile, 2005), and it is likely that some reports of research that could have been missed. Additionally, some inquiries narrated more exhaustive evidence than others, making it simpler to cater to specific illustrations of EL, non-EL achievement gaps, teacher education, and infusion model. Nonetheless, this literature review has blended an adequate body of research to bolster the implications afforded in the discussion and to perform as a base for forthcoming researchers desiring to give these issues a mainstream outlook.

Achievement Gap, Reasons, Consequences, and Solutions

The term 'achievement gap' refers to inequality in educational outcomes between *White* students and students of other ethnic, linguistic, and racial background (Banks, 2012; Anderson et al., 2007). In other words, the achievement gaps result when a group of students (e.g., students categorized by race/ethnicity, gender, socioeconomic status, type of school, etc.) performs better than another group and the disparity in mean scores for the two groups is statistically significant, i.e., exceeds the margin of errors (National Center for Educational Statistics). Achievement gaps can be of two types, (a) internal gaps, i.e., the average difference between distinct ethnic, racial, and linguistic groups and their *White* counterparts within a school, and (b) external gaps, i.e., the mean difference between the total school scores for each student subgroup in the school and average group scores for *White* students within the state (Anderson et al., 2007). The federal government in the United States started keeping educational records of educational outcomes of three major national, ethnic groups, *Non-Hispanic Whites*, *Black*, and *Hispanic*, in 1971 (Portes, 2005). Currently, achievement gap data related to a wide range of student groups are available, which have been instrumental in understanding their underlying characteristics.

A wealth of evidence makes it clear that there is as big as one standard deviation learning gap between African American and White, and Hispanic and White K-12 student population

(Banks, 2012). Similar extrapolation can be made for the discrepancies between English language learners and non-ELs (Nutta et al., 2012; Welch-Ross, 2010). These assertions are true and evident until the day of this review. The academic gap is evident in almost every measure of educational progress including, grade point average, standardized tests, the dropout rate, and the degree to which students are left back a grade. The achievement gap between minority and economically underprivileged students and their more affluent peers endures despite the tireless efforts by countless educators, massive education restructuring initiatives, numerous state and local programs, grants from the federal government and scores of institutes engaged in educational advancement, and an enormous body of educational theory, research, and annotation (Troff, 2014). Thus, many researchers believe that achievement gap is a complex issue and should be studied from multiple perspectives such as motivation (Cooper & Schleser, 2006); age (Sanders et al., 2018; Banks, 2012), time, geography (Banks, 2012) teacher education (Yoder, Kibler, & van Hover, 2016; Darling-Hammond, 2004; Darling-Hammond, 2014b; Good, Maswwicz, & Vogel, 2010; Karbenick & Noda, 2004), school and classroom environment (Damme, Fraine, Landeghem, Opdenakker, & Onghena, 2002; Berkowitz, Moore, Astor, & Benbenishty, 2017), class size (Ferguson, 2003), technology, motivation, learning environment, and curriculum (Schofield, 2010; Lee, Maerten-Rivera, Penfield, LeRoy, & Secada, 2008; Shcuetz & Schuetz, 2005), assessment (Abedi, 2002; Hung et al., 2019), opportunity gap (Hung et al., 2019), Communication gaps between teacher and student, teacher and parents, parents and children, and school and school, lack of support for families in transition to a new environment and culture, and Lack of systemic articulated district EL Plan (Good, Masewicz, & Vogel, 2010), educational policy and procedures (Portes, 2005; Grissmer, Treisman, Fuller, & Meyer, 2001).

As mentioned in Kim (2011), EL status and race among K-12 students were statistically significant predictors of school dropout. English learners were approximately two times more likely to drop out compared to non-ELs. Likewise, Hispanic and Black students were 2.6 and 1.7 times more likely to drop out from their schools compared to Whites. Hence, the chances of reducing dropouts are directly correlated to students' academic achievement. The researcher found that the odds of dropping out decreased by nearly half in reading, and more than half in math for students whose test scores are one Standard Deviation higher. The most important conclusion of this research was, 'in terms of demographics, the largest differences between ELs and non-EL students exist in ethnic composition and FRL status' (Kim, 2011, p. 46).

One of the most recent studies in this area investigated the impact of students' mobility and segregation on ever-increasing gaps. Using a 6-year panel of student-level data (2007/08 through 2012/13) from Clark County School District (CCSD) who were enrolled for at least two consecutive academic years, Welsh looked into 428,247 students based on their gender, race/ethnicity, *FRPL*, ELs, and special education status, and their performance on reading and math in Grades 3 through 8, & High School Proficiency Exam (HSPE) in Grade 10 (Welsh, 2018, p. 64-65). The results showed that high within-year student mobility rates were directly correlated to low-achieving, high-minority, and high poverty schools. A high correlation between low achieving schools and the number of disciplinary actions suggested that the low-achieving student population possesses greater behavioral management challenges. As mentioned by Welsh (2018), schools grouped as 'in need of improvement' had the most elevated student mobility rates, while 'high-achieving and above' schools had the lowest midyear withdrawal rates (p. 71).

All in all, Welsh (2018) concluded that achievement gap based on income/SES remained constant for 6-years, while the differences based on ethnicity had mixed phenomena during the period of study. The movement of low-achieving minority students, in most cases, was either for other low achieving school or even lower, for example, the researcher found that 75% of midyear movers and 80% of between-year movers were transferred to other low-achieving schools. These results are highly comparable with the results of similar studies that focused specific state or school districts, like Madison Metropolitan School District by Broman, Grigg, & Hanselman (2016); Rural Idaho Schools (Brown, 2017); the State of Indiana (Faith, 2017); Northern California & Wisconsin (Herman, 2009); the State of Texas (Linton & Kester, 2003; Hanushek & Rivkin, 2006), etc.

Jeynes (2015) presented the results of a meta-analysis of 30 prior quantitative studies conducted between 1975 through 2009, which examined the relationship between factors that various programs have utilized to intervene to reduce the racial achievement gap and the size of such reduction. Total of 491,197 students were included in these research studies to assess the overall effectiveness of programs designed to reduce the achievement gap and the finding suggested that effect sizes for ages, standardized and unstandardized achievement measures, classroom structures, cultural factors, and high expectations in religious faith were in the positive direction, but they were not statistically significant. Conversely, the factors like curriculum, family factors, and religiously oriented school programs had been able to lower the achievement gaps statistically significantly; however, the government policies during the period of studies included in this meta-analysis were found to have a negative contribution in this regard (Jeynes, 2015).

A survey conducted by Uro & Barrio in the year 2013 on a total of 65 schools in the member districts of the Council of the Great City Schools found that ELs failed to keep pace with their non-ELs counterpart in terms of academic achievement. A close look into the underlying factors shed light on the issues behind this level of disparity among student groups. The ELs were being taught in mainstream classrooms by general teachers who did not meet the professional development or certification requirements to help these students. Most of the schools in the member school districts did not offer EL specific professional development opportunities to general teachers, which restricted achievement growth among low achieving student groups (Uro & Barrio, 2013).

Along the same lines, despite the widely accepted belief that the language of mathematics is universal and free of cultural influences (e.g., Hoffert, 2009; Meaney & Flett, 2006), the learning gap between EL and non-ELs is as big as in any other subjects taught in mainstream classrooms. Many researchers reported that language fluency is directly associated with students' accomplishment in content area courses including mathematics (Cobb & McClain, 2004; Kazima & Adler, 2006; Anthony & Walshaw, 2007; Boero, Douek, & Ferrari, 2008; Xi & Yeping, 2008) and to eliminate confusion and challenges especially scaffolding prior knowledge during everyday classroom (National Council of Teachers of Mathematics, 2000; Boero et al., 2008; Lavy & Mashiach-Eizenberg, 2009; Barwell, 2005).

A close look in the results of many empirical studies, including various longitudinal studies, an ominous picture of achievement gaps is consistently threatening our educational system, despite multiple attempts to reduce them. It has been clear that the gap is not dormant, but dynamic and susceptible to other academic features, which further expands the divide. It can accelerate negative tendencies like segregation (Welsh, 2018); use of remedial curriculum (Beecher & Sweeny, 2008) which prevents students from exercising higher cognitive functions (Cooper & Schleser, 2006); stereotype threats that overwhelmingly affect low-income, low-achieving students (Borman et al., 2005; Furgione et al., 2018); inequitable treatment (Carrell & Dittrich, 1978); larger class size and inequities that potentially undermine learning (Darling-Hammond, 2015); differential rate of discipline infractions (Faith, 2017); confirmation of racial hierarchy (Herman, 2009); differential development of executive function skills (i.e., working memory, cognitive flexibility, and inhibitory control) based on student groups (Little, 2017).

There has been a plethora of research on reducing the achievement gap, but there is a lack of consensus about the strategies the U.S. school system requires at various levels to realize the gap abate (Jeynes, 2015). None of the low achieving student groups, including ELs, have easy access to resources to have better outcomes in standardized testing (Hunter & Bartee, 2003; Jeynes, 2015) even after the educational reform initiative known as the No Child Left Behind Act of (2001). Recent conceptualizations of access assert that it should be evidenced in students' academic outcomes (Gilmour, Fuchs, & Wehby, 2019).

Education is seen as the great equalizer in the face of structural inequalities, but it often falls short along the line of racial, ethnic, linguistic, and socioeconomic status (Hung et al., 2019). In this context, it is important to realize that ELs bring various complex experiences like poverty, political oppression, and interrupted schooling (Villegas, de la Mora, Martin, & Mills, 2018) in their everyday classrooms. For a long time, these students were taught by teachers with specialized language and pedagogical skills until ELs could participate in instruction in English. As noted by Karbenick & Noda (2004), most in-service teachers lacked basic foundational knowledge about EL issues, even if 88% of them had taught or were teaching ELs in their classrooms. It explains the reason why the mainstream teachers were woefully unprepared for reaching out to all the students (McHatton & Parker, 2013; Coady, Harper & de Jong, 2015) or successfully running all-inclusive classrooms. As Troff (2014) asserted, teachers and educators with little to no experience of low achieving linguistic minority students tend to believe that ELs lack lower-order fundamentals, thus, they are unable to benefit from higher-order challenges in their classes and tend to assign remedial curricula that emphasize drilling the fundamentals, which further lowers challenges, thereby dropping academic achievement further down. Thus, the continuation of the EL and non-EL achievement gap prompted urgent actions from the U.S. Department of Education in terms of arming in-service and preservice teachers with required professional knowledge and credentials by adding U.S. federal obligations aiming for teacher quality and student accountability (U. S. Department of Education, 2011).

Successful inclusion is dependent mainly on teachers' attitudes toward the students with disabilities and minority backgrounds, and such positions can be positively instilled, inspired, and/or confirmed (de Boer, Pijl, & Minnaert, 2011; Forlin, 2010). Teachers with positive attitudes are believed to be more willing to work with students will diverse cognitive, ethnic, racial, and linguistic needs in general mainstream classrooms (Klingner & Artiles, 2003; Rheams & Bain, 2005). On the other hand, many educators and second language acquisition (SLA) researchers also believe that the existing content area assessments are unfair because they tend to test students' knowledge of the English language rather than the subject-specific knowledge (O'Conner, Abedi, & Tung, 2012).

Thus, the pursuit of effective teaching and quality learning within the realm of accountability is currently being promoted as the means of mitigating academic achievement gaps between the historically low-achieving student and their high-achieving counterparts. It is believed that quality input accelerates output among students strong enough to narrow the achievement gaps (Hunter & Bartee, 2003). Compared to the amount of debate over the achievement gap in the last few decades, the volume of quantitative research is thin, and it is important to realize that such studies must guide our way forward to come to any definitive conclusions about achievement gaps (Jeynes, 2015).

Admittedly, the findings of the studies included above clearly establish that the achievement gap is not a new phenomenon. Students' success in the existing testing system seems to be connected to ethnicity, English language status, gender, disability status, and socio-economic status, even if we keep every other variable (e.g., parents' education, etc.) constant. The gaps in many cases are either as big as they were a decade ago, if not bigger (Abedi, 2002; Banks, 2012; Darling-Hammond, 2004). It is pressing that educators, researchers, and policymakers take immediate reasonable actions toward finding the reasons behind the unfair practices which propel these gaps.

Teachers' Impact on Students' Learning Outcomes

Empirical studies in the field of teachers' impact on students' learning outcomes have overwhelmingly supported the idea that teachers are highly instrumental in bridging the achievement gap (Johnson, & Wells, 2017; Darling-Hammond, 2015; Darling-Hammond, 2002; Chetty, Friedman, & Rockoff, 2014). The quality of schools depends on the quality of teachers.

More effective learning on the part of the student ties to the learning opportunities their teachers receive during their teacher preparation years (Feiman-Nemser, 2001). Policymakers and educators believe that "what students learn is directly related to what and how teachers teach; and what and how teachers teach depends on the knowledge, skills, and commitments they bring to their teaching opportunities, they have to continue learning in and from their practice" (Feiman-Nemser, 2001, p. 1013). Positive correlations have been reported for a long time between teachers' intelligence or academic ability, subject matter knowledge, pedagogical knowledge, teaching experience, and certification status, and students' learning outcomes (Darling-Hammond, 2000; Banks, 2012).

One of the seminal works in this area entitled '*Teacher quality and student achievement: A review of state policy evidence*' by Darling-Hammond (2000) directly examined how teacher qualifications and other school inputs were related to student achievement. Using 1993-94 Schools and Staffing Surveys (SASS), and the NAEP survey data, this mixed-method study shed light on two major fronts, (a) measures of teacher preparation and certification were the strongest correlates of student achievement in reading and mathematics, both before and after controlling for students' SES and language status; and (b) policies adopted by states regarding teacher education, licensing, hiring, and professional development would make an essential difference in the qualification and capacities that teachers bring to their work (Darling-Hammond, 2000, p. 1).

Findings of this study suggested a few noteworthy conclusions:

- student demographic characteristics were strongly correlated to student learning outcomes at the state level; however, they were less influential in predicting achievement levels compared to the variables used to measure teacher qualities.

- the average state-level teacher quality variables correlated more strongly to student achievement than class sizes, schools' expenses on academic resources, and teacher salaries.
- among teacher quality measures, teachers' major, and their full certification status, were more powerful predictors of student achievement than teachers' education levels; and
- changes in course taking, curriculum content, testing or textbooks were less influential in student outcomes compared to the teacher qualities variables, i.e., teachers without proper knowledge and training did not know how to diagnose their students' learning needs and how to use available tools to accomplish desired goals (Darling-Hammond, 2000, p. 31-33).

These findings are comparable to most of the recent studies that looked into the effect of teacher qualities on student outcomes from multiple perspectives, e.g., pedagogical knowledge practices in STEM education for elementary schools (Hudson, Lane, & Pullen, 2005); integration of information and communication technologies (Koh, Chai, & Lim, 2017); teachers' emotional exhaustion (Arens & Morin, 2016); influences on policy and practices (Darling-Hammond, 2014a); student-teacher relationship (McGrath & Bergen, 2015); teacher self-efficacy and wellbeing (Zee & Koomen, 2016); teacher clinical practices (Burn, Mutton, Thompson, Ingram, McNicholl, & Firth, 2016); teacher servant leadership (Noland & Richards, 2015); teachers' classroom management skills (Korpershoek, Harms, de Boer, can Kuijk, & Doolaard, 2016; Akalin & Sucuoglu, 2015) and many others.

McHatton and Parker (2013) conducted a longitudinal study that compared elementary and special education preservice teachers' perceptions and attitudes towards inclusive classrooms and its significance on their attitude towards inclusive practices vis-à-vis the courses the PSTs take during their studies. The participants (N = 56) were university PSTs who attended a large public university in the southeastern United States and completed at least a year into their programs. Of total participants, 31 (n_1 = 31) were elementary education majors, while 25 (n_2 = 25) were special education majors. Despite being under the same department, the special education majors had small class sizes (25 per class), and the program progressed through sequenced courses, unlike the large class sizes (maximum of 300), flexible and unstructured string of courses of elementary education majors. Also, the special education track exposed the PSTs more towards the field of disabilities and the EL issues, and an emphasis on consultation, collaboration, and problem-solving at the preservice level compared to elementary education majors (McHatton, & Parker, 2013).

The researchers measured the PSTs' pre- and post-treatment attitude towards inclusion of all students using the 'Attitude Toward Inclusion survey' that helped them identify the current understanding of PSTs' notion of integration. The survey used a 5-point Likert Scale ranging from 1 through 5 (1 being strongly disagree, and 5 being strongly agree). They conducted a delayed posttest after one year of the treatment to better assess the outcome of the procedure. Finally, they ran a repeated-measures ANOVA using PSTs' major as between-subject factor and time of testing as within-subject factors. The treatment included a series of activities conducted under the supervision of two faculty members from the Department of Special Education and one from Childhood Education. The participants engaged in co-teaching activities throughout a semester in classroom management/instructional method courses specific to special-needs students, the course was systematically linked to a field experience, and they spent time co-teaching their peers who were monitored by in-service teachers. During this period, the participants got various opportunities to discuss and review what they did and learned through these processes (McHatton & Parker, 2013).

The findings suggested that the elementary education PSTs had less positive attitude mean scores ($\underline{M}$=3.89) towards all-inclusive class compared to special education PSTs ($\underline{M1}$=4.4). The mean attitude scores increased for both groups toward the end of the study period after a yearlong intervention; however, the mean difference persisted (*EMS* = 4.16; *SEM* = 4.39). There were statistically significantly higher means overall for SEMs $F_{(1, 54)} = 5.74$, $p = .001$ at the beginning and the end of the study period. There were statistically significant differences for both SEMs and EEMs from Time 1, $F_{(1, 54)} = 15.51$, $p < .001$, and Time 2, $F_{(1, 54)} = 4.41$, $p =$ .040. However, the rate of change in attitude was significantly higher for EEMs than for SEMs. Furthermore, there was a statistically significant interaction effect between the time and PSTs' area of study (McHatton & Parker, 2013).

Furthermore, the results of the delayed posttest revealed that there did not exist a statistically significant difference compared to Time 2, $F_{(1, 52)} = .689$, $p = .41$, and there was no statistically significant interaction effect between time and the major, $F_{(1, 52)} = 0.051$, $p \geq .05$. That means the groups were comparable at the end of the study period and a year after that. In addition, the descriptive analysis of the survey data collected in three different periods revealed that 56% of SEMs and 54.8% of EEMs strongly agreed that inclusion was beneficial for all students, and 100% of SEMs and 90.3% of EEMs strongly agreed or agreed that inclusion raised understanding and acceptance of differences. These ratings went up (90.3 % - 100%) and remained comparable to posttest ratings even during the delayed posttest period. However, the rate of 68% of SEMs and 29% of EEMs who strongly agreed or agreed that including students with disabilities would not negatively impact school performance on high-stakes tests, increased to 61.3% for EEMs but decreased to 58.3% for SEMs (McHatton & Parker, 2013).

Based on the above findings, the researchers concluded that elementary and special education PSTs follow distinctly diverse trajectories in acquiring attitudes toward inclusion across their TPPs, e.g., disconnect in attitude towards inclusion of students with exceptionalities. The content area (math, science, social studies, etc.) PSTs welcome opportunities to learn the theories, skills, and like to participate in hands-on training to prepare themselves to reach out to all students, and they have a higher rate of positive changes over time, compared to SEMs that possessed higher overall mean scores at the beginning and remained remarkably stagnated over time (McHatton & Parker, 2013).

A study conducted by Kindler (2003) that considered the data from 41 state education agencies in the United States reported that only 18.7% of ELs met state norms for reading in English (As cited in Genesee, Lindholm-Leary, Saunders, & Christian, 2009). Students who come from linguistic minority families have higher dropout rates and are placed in lower ability groups than their non-EL counterparts (Genesse et al., 2009). Today, K-12 students come from diverse minority households that speak more than 400 home languages. The population of ELs doubled within the last few decades, while the total student enrollment roughly increased by 12%

within this period. More than 4.8 million school-age children were identified as ELs in the year 2018, which was 9.5% of the K-12 public school student population (U. S. Department of Education). This significant shift in society calls for high standards and strong accountability for schools, students, and families. The reauthorization of the Elementary and Secondary Education Act 2001, No Child Left Behind (NCLB), included ELs in the state accountability system and demanded students be tested in reading and mathematics annually. In addition to that, ELs were required to take English language proficiency tests every year. In this context, an enhanced education is the only way to increase ELs' performance on such high-stakes tests and to reduce the achievement gap (Genesse et al., 2009). This value-added measure of teachers' assessment is the primary factor in determining overall school funding resources, and their finances, and professional goals (Portes, 2005).

Current educational reform concerning the achievement gap sheds an intense focus on teacher quality (Townsend & Bates, 2007). It necessitates that TPPs design, implement, and evaluate curriculum, field experiences, and clinical practices so that PSTs acquire the knowledge, skills, and disposition required to facilitate all students learn (Cochran-Smith, 2001).

One of the pivotal works in this area was conducted by Loeb and the team that sought answers for three research questions, (a) how much does teacher effectiveness vary across classrooms for EL and non-EL students? (b) are teachers equally effective with ELs and non-ELs, and (c) can measures of teacher characteristics help explain differences in teacher effectiveness? (Loeb, et al., 2014, p. 458). Using data from the Miami-Dade County public schools from 2004/05 through 2010/11, they intended to gauge the differential teacher effectiveness with ELs using Value-Added Model (VAM) of teacher effectiveness, which calculated "value-added estimates in the form of coefficients on teacher fixed effects used to predict student test score gains" (Loeb et al., 2014, p. 461). Of over 347,000 students (N > 347,000), more than 225,000 were *Hispanic* and approximately 67,000 English learners.

They used multiple statistical procedures, for example, a bootstrapping approach to answer the first research question; cross-tabulation, the parametric and non-parametric correlation for the second research question; and multivariate regression analysis for the third research question. The findings suggested that (a) except in high school math, there was no significant difference in the actual variance in value-added of teacher effects for EL and non-ELs, (b) teachers who were good with ELs tend to be good with non-ELs and vice versa, and (c) teachers who were fluent in students' native language had a statistically significantly higher positive impact on students' learning gains (Loeb et al., 2014).

Teachers need to know three facets of subject matter knowledge: (a) familiarity of vital facts, concepts, theories, and procedures on their field; (b) knowledge of illustrative outlines that shape and link ideas; and (c) familiarity of the rules of substantiation and proof (Shulman, 1986; Feiman-Nemser, 2001). The primary duty of a teacher is to help their students learn meaningful content, which requires a thorough understanding of the subject they teach. Likewise, teachers also need to know their subject from a pedagogical point of view, which entails knowing what students find puzzling or problematic and having substitution descriptions, models, and analogies by framing well-thought-out purposes about the content in advance through well-designed curricular materials (Feiman-Nemser, 2001).

Although there isn't enough evidence to conclude that the body of knowledge and skills a teacher puts on the table in everyday classroom boosts their students' learning gains (Coady, Harper, & de Jong, 2015), it is widely believed that they get automatically translated into effective instructional practices, which ensures the equity in academic achievement (Kane & Staiger, 2002; Ho & Kane, 2013). There are several empirical studies (e.g., Darling-Hammond, 2000; Wilson, Floden, & Ferrini-Mundy, 2001; Darling-Hammond et al., 2005; Aaronson, Barrow, & Sander, 2007; Koedel & Betts, 2007; McCaffrey et al., 2009; Measuring Teachers' Effectiveness (MET) Project, 2009-2012; Baker et al., 2010; Glazerman et al., 2010) that hint toward the close cohesiveness of teacher preparation programs, and instructional practices to students' learning outcomes. Standing on the foundation of this research, we can assert that teacher education is not the single most important aspect to ensure better academic achievement, but it is, indeed, one of the crucial elements to ensure equitable teaching practices.

Teacher Education and Attempts to Narrow EL, Non-EL Achievement Gaps

As mentioned earlier, teachers in the United States increasingly work with a diverse and growing set of students, including ELs whose learning needs are unique (Coady, Harper, & de Jong, 2015). Since the 1990s, the U. S. K-12 system has started to change due to three underlying reasons, (a) the enrollment rates of ELs overtook the capacity of bilingual and ESL programs, (b) the political backlash against bilingual education in some states like California, Arizona, etc. and, (c) the implementation of NCLB act of 2001 which delegated that ELs be included in states' testing programs and their scores reported as a subgroup followed by federal accountability measures, the Common Core State Standards. It has forced schools to mainstream ELs with the belief that English Immersion programs boost their test scores (Villegas et al., 2018). The following quote from Nutta et al. (2015) highlights new opportunities and challenges for ELs and their teachers:

> *"On one hand, the standards present EL students with an opportunity to gain equal access to rigorous instruction along with higher expectations for learning.*
> *On the other hand, teachers will need to make challenging new standards accessible to English learners, regardless of their English proficiency. These opportunities and challenges will, in turn, require a change in how teachers view and implement instruction for English learners in the mainstream classroom. Educators in every state of our nation are confronting the challenges English learners face in accessing more rigorous instruction, a gap they need to address now more than ever."* (p. 1)

Given the high accountability placed on teachers, established evidence of teachers' effectiveness in learning gains, steadily increasing EL population, and the lingering problem of the achievement gap, many university-based teacher education programs have implemented a litany of instructional initiatives to better prepare teachers to work with students from diverse backgrounds beginning in the mid-70s (Banks, 2012; Townsend, & Bates, 2007). If regular standards highlight teaching as "telling and learning as listening," progressive archetypes seek teachers to do more listening as they prompt student thinking and appraise their belief and "for students to do more asking and explaining as they investigate authentic problems and share their solutions" (Feiman-Nemser, 2001). It requires teachers to be practical thinkers, curriculum developers, and creators of knowledge in praxis, which requires teachers to work harder in general for educational change (Feiman-Nemser, 2001). The *Encyclopedia of Multicultural Education argues that*:

There is a consensus among educators, researchers, policymakers, and the general public that despite the existence of some outstanding teacher education programs, the education of teachers in the United States needs to become much better and do more to contribute to addressing the opportunity and achievement gaps that continue to exist in public education.

(Banks, 2012, p. 2118)

This proposition to reform education, curriculum, pedagogy, and areas of student welfare and support, are mostly driven by school administrators and teachers in an attempt to seek improvement on what they do on a daily basis, and how can they boost the learning gains among students regardless of stereotypical characteristics assigned to existing gaps (Townsend, & Bates, 2007).

Research in TPPs often considers isolated effective practices such as field-based teaching, case-based teaching, portfolio assessment, and even student or faculty cohorts, etc. and their impact. It is essential to dissect such programs and take a close look into individual components, but the effectiveness and meaning of such isolated practices always depend on the overall objectives they serve. For example, studies in the field of Science, Technology, Engineering, & Math (STEM) teacher education have sequentially identified some essential elements that TPPs require to integrate into their curriculum, e.g., field experience and community-based experiences (Wilson et al., 2001) to provide opportunities for PSTs to gain first-hand experience of teaching and to see how the theoretical classroom concepts translate in reality (Francis & Stephens, 2018); and to collaborate with diverse ethnolinguistic communities to make self-exploration of strategies and methodologies about teaching STEM courses to ELs

(Garcia et al., 2010). Gross and the team (2010) underscored the value of PSTs' self-awareness of their perception as teachers and ELs during the field experience. Researchers have reported that multiple field visits enable PSTs to acknowledge their teacher identities more precisely and appreciate their maturity over time (Francis & Stephens, 2018; Gross, Fitts, Goodson-Espy, & Clark, 2010). The field experiences widely vary by the institutions, and though there are a few reports of field experiences with ELs that were not tied with the goals and the content of the TPPs, many studies have reported that PSTs' field experiences are the most critical component in providing them with a high degree of preparedness and efficacy to work with ELs (Wilson et al., 2001).

After a lengthy discussion about teaching to learn, researchers have realized that the education in its entirety is greater than the sum of its parts, which requires an in-depth study of the TPPs in general, rather than its single aspect. A conceptual outline of TPP should explicitly lay out the definite themes, view of learning, the role of teacher educators and PSTs, and the goal of schooling (Howey, & Zimpher, 1989) to enhance the likelihood of getting preferred outcomes by instilling fundamental skills in PSTs to shape curriculum, culture, pedagogy, and assessment practices. A Purposeful design of field experience (e.g., classroom observations, internship, guided practice, knowledge application, and inquiry) and its proper use are other important indicators of program coherence. A well-thought-out sequence of multiple placements enables PSTs to observe and practice teaching they learn in their classes as they shift from observation to restricted partaking to full accountability with suitable modeling and supervision. Exemplary TPPs aid PSTs' learning long after they graduate by creating a stimulating culture and a coherent curriculum, by examining their responses to changing concepts and practices, and by offering required assistance and challenge their changing knowledge, skills, and beliefs (Feiman-Nemser, 2001).

Consigning meaningful and persistent teacher education at the epicenter of school reform was a profound idea during the 90s. The TPPs should allow and prepare PSTs to establish and advance rudimentary inventory for good teaching, which entails specific characteristics not limited to lead discussions, plan experiments, arrange parent-teacher and student-teacher conferences, set up classroom libraries, organize and execute writer's workshop, and arrange field trips, etc. TPPs should give PSTs an insight to formulate customs and skills compulsory for the ongoing study of teaching in the company of colleagues. PSTs must come to see that learning is a fundamental part of teaching and that critical dialogues about teaching are cherished resources in developing and refining their procedure. In general, professional preparation for PSTs comprises courses in educational foundations and general and/or specific methods of teaching. Educational psychology is a staple in educational foundations, but courses in philosophy or history have been replaced with an 'introduction to teaching' course. All programs require some supervised practice called student teaching. Dispersed courses taught under different departments seldom build on or bond to one another, nor do they come together as a consistent preparation for teaching. Without a set of organizing premises, shared principles, and evident goals for PSTs, there is no basis to pilot program design or student evaluation (Feiman-Nemser, 2001).

In this context, especially over the past decade, there have been some serious endeavors to identify and tackle the challenge of preparing mainstream teachers for linguistically diverse classrooms, however, this body of research is small but quickly expanding. It has not been synthesized enough, which deprives teacher educators of understandings what it takes to prepare mainstream teachers for their new role as teachers of ELs (Villegas et al., 2018). Furthermore, some scholars and researchers started to put together the knowledge base, and skills a teacher of ELs required to teach both content-area and language equally effectively at the same time (Coady, Harper, & de Jong, 2015; Wong-Fillmore & Snow, 2000). These premises highlight the importance of both linguistic and cultural dimensions of schooling (Coady & Escamilla, 2005) to simultaneously facilitate ELs' content-area knowledge and linguistic development (de Jong, Harper, & Coady, 2013; Lucas & Villegas, 2008).

The knowledge and beliefs that one brings into their PST classrooms have a profound influence on what they learn (Villegas et al., 2018). Prior convictions sometimes barricade the propensity to embrace reforms if not guided and checked by an expert vision. Such prior beliefs come from their long experience as students in elementary, middle, and high schools, including subject matter, and the images of teaching, learning in general. Such beliefs might potentially misinform PSTs that they know more about educating than they genuinely do, which prevents them from being adaptable towards new ideas, actions, and thoughts. The TPPs must give enough opportunities for their PSTs to critically examine their deep-rooted beliefs and to decide whether to confirm or amend them (Feiman-Nemser, 2001).

TPPs should be guided by the idea that PSTs require a lot of experience of teaching through different stages of their learning to teach. The robust learning environment the PSTs receive through child study, classroom inquiry, co-planning, co-teaching, and other forms of assisted classroom endeavors before they walk in as independent teachers positively impact their attitudes towards teaching in general. Thus, Feiman-Nemser (2001) concluded that the central task of TPPs is to assemble existing rationales about what teachers need to know and be able to progress extensive learning for all students.

To help students make meaningful connections to the subject matter, teachers need to know about their students, including their development and the process of learning (Villegas et al., 2018). Most of the studies in this meta-analysis analyzed opportunities that the PSTs received in their TPPs' to knowing ELs and their difficulties while attending classes in English. Many of them noted that (a) ELs' proficiency of L2 is directly associated with the level of challenges; (b) their knowledge of Basic Interpersonal Communication Skills (BICS) does not guarantee academic success; and (c) familiarity of linguistic principle could help PSTs realize that the academic complications many ELs experience in schools were likely to be a usual part of language development and they can facilitate this journey regardless of their area of teaching (Villegas et al., 2018).

Approximately 3 out of 4 studies included in meta-analysis addressed teachers' beliefs about ELLs. Thirteen of sixteen empirical research discussed PSTs' views about teaching ELs. It is crucial for the TPPs to promote PSTs' openness toward the inclusion of ELs in mainstream classrooms and design programs to embed components (e.g., field experiences that enable direct contact with ELs, observing ELs in mainstream classrooms, working with ELs in their homes, and assisting ELs in after-school programs) that help explore these areas. Twelve studies addressed PSTs' views of ELs, and in 3 of them, the PSTs' substantially broadened their entering views of ELs, which promoted PSTs' sense of empathy, sympathy, and appreciation for ELs. The learning opportunities were purposefully designed. PSTs need to develop a deep understanding of the concepts in the discipline of their interest and make connections with other, e.g., modes of reasoning—an understanding of the nature of knowledge and represent them to students. Four of twenty-one studies directly addressed the development of subject matters for teaching, and they found that ELs in mainstream classrooms faced dual challenges of learning content and language. The findings suggested that PSTs should be taught the effective ways of mastering their subject matter, also, the use of language in them so that they can successfully scaffold ELs' learning of content and language (Villegas et al., 2018).

To prepare PSTs to make informed academic choices once they enter the vocation, teacher preparation programs (TPPs) must help them develop a beginning repertoire of ways to engage learners effectively, which involves developing PST' familiarity and skills with well-designed curricular materials, sound instructional methods, and accurate models of assessment that support student learning. Nine of twenty-one studies examined PSTs' learning prospects envisioned to develop some aspects of teacher candidates' beginning repertoire for teaching ELs especially by embedding components that help them to scaffold ELs' knowledge accurately, to create welcoming and safe classrooms that encourage ELs to take risks and to help them design equitable assessments for ELs by reducing linguistic complexity. They found that the PSTs were more inclined to practice different scaffolding techniques to help ELs meet rigorous academic objectives; they had a significantly higher level of commitment to creating a collaborative classroom that makes ELs feel comfortable to speak in English; and PSTs conceptions of assessment expanded substantially and learned multiple strategies to make assessment equitable for ELs (Villegas et al., 2018).

As part of their preservice preparation, aspiring teachers must develop intellectual tools to study teaching, which entails PSTs' observational skills, interpretation, analytical skills, and reflection. An overwhelming amount of research in this area found inquiry and reflection to be the two most significant practices that TPPs need to teach PSTs to enable them to study and enrich their teaching meticulously (Villegas et al., 2018; Athanases et al., 2013; Settlage et al., 2014). Six of twenty-one studies that focused on the inquiry processes which involved the activities like, (a) finding patterns in student data, documenting student achievement, (b) analyzing student work (Athanases et al., 2013), (c) developing an EL case study and helping them figure out specific teaching strategy (Jimenez-Silva, & Olson, 2012), (d) writing narrative about ELs (Pu, 2012), and (e) observing learners in various classrooms to identify teaching practices (Virtue, 2009) supported the findings that PSTs were better able to make connections between theory and practices (Villegas et al., 2018). Eighteen of twenty-one studies in this meta-analysis looked into the effect of PSTs' reflection on sociocultural learning by letting them reflect on their experience as students in their L2 classes in which teachers used other than English language and model scaffolding ELs' learning (Galguera, 2011; Settlage et al., 2014); as classroom observers (Hutchinson, 2013); as study abroad participants (Medina et al., 2015); engage in autobiographical writing (Markos, 2012); write reflective journal over time to observe the evolution of their knowledge about ELs (Siegel, 2014); or participate in reflective problem-solving discussions related to a teaching case involving ELs (Wade et al., 2008).

Such opportunities put PSTs through a rigorous process of revisiting their own experiences, which helped them be better organized for their future teaching and positioning learned ideas within specific tasks enabling them to understand theories in higher terms and finding better ways to practice them (Villegas et al., 2018).

Many present-day amendments in TPPs call for content-rich, learner-centered teaching, which stresses intangible acumen and offers all students' chances to think judiciously, solve problems, and absorb things that matter to them and have significance in the outside world (Feiman-Nemser, 2001). Based on various reported empirical research and findings, Johnson & Wells (2017) recommended that teacher education programs in the United States include three key components and there is enough subsequent literature to strengthen their plausibility, (a) field-work for preservice teachers (Nutta et al., 2012), (b) professional development on pedagogical language knowledge and multicultural sensitivity training (Karabenick & Noda, 2004; Walker, Shafere & Iams, 2004), and (c) a holistic, and aligned evaluation process to strengthen new English Language Arts (ELA) teachers' chances of meeting the linguistic need of English learners (Byrnes, Kiger, & Manning, 1997). There is an intense debate about the kind of teaching and teacher education that should delineate instruction in the twenty-first century, concerning the most basic questions like who should prepare teachers, when and where this preparation should take place, and what should be the content of the preparation program.

Historically, teacher education involved defining and discovering the best approaches built on various assumptions and beliefs for public education. Currently, the discussion has evolved into two distinct schools of thought. The first one strives to prepare professional and highly qualified teaching force through teacher education (Zeichner, 2014; Darling-Hammond, 2005), while the second promotes priming teachers as technicians and letting them implement the teaching scripts (Zeichner, 2014). The U.S. has been dealing with the severe gap of opportunities to learn, school completion rates, and academic achievement, including inequalities in access to resources and qualified and dedicated teaching forces for different segments of the population (Zeicher 2014; Duncan and Murnane, 2011). The inequities have at least upheld if not widened the gaps among student groups in K-12 public schools. The existing teacher education system in the U.S. has not been able to supply the most poverty-impacted schools with more experienced teachers to help them breakthrough their dependency on inexperienced and underprepared teachers (Zeicher 2014; NCTAF 2010).

The U.S. government and the private educational foundations have initiated two initial responses against lingering problems of teacher education for the last four decades. The first initiative started during the mid-60s, which involved efforts to build an effective system of teacher education within universities. The federal government invested a considerable amount of money in the form of competitive grants administered and monitored either from Washington D.C. or through state education departments. Moreover, many private foundations that include, Carnegie Corporation, the Ford Foundation, and the Bill and Melinda Gates Foundation have put a substantial amount of money in research and supplies targeting high poverty communities (Zeicher, 2014; Sykes, & Dibner, 2009).

Second, in reaction against 'high-cost, low technology use, and less effective' university-based teacher education, some critics believe in creating a competitive market for teacher education (Furlong, Cochran-Smith, & Brennan, 2013; Klees, 2008). They view university-based teacher education as a barrier for prospective teachers to enter this field (Cochran-Smith & Villegas, 2015). This idea has recently gained momentum, for example, many regulations have been cut during Bush and Obama administrations, and influential initiatives like 'Innovation in Education' have recently awarded 263 million dollars to the organizations like 'Teach for

America,' 'New Teacher Project,' 'Boston Teacher Residency Program,' etc. (Zeichner, 2014). However, both reactions agree on the point that teacher quality should be drastically enhanced.

PSTs need to learn to develop an academic posture engrained in the knowledge of child/adolescent development and learning. That means, be mindful of the distinguishing characteristics and capabilities of students of different ages, and their unique ways of thinking and acting molded by their language and cultures. Conscious perceptions on development and learning afford essential contexts for understanding students, creating proper learning activities, rationalizing didactic choices and actions, and working with parents, students, administrators, and colleagues (Feiman-Nemser, 2001).

The fragile relationship between courses taught to the PSTs and clinical/field experiences portrays a lack of coherence in TPPs. It is important to remember that PSTs deem firsthand teaching as the most prized part of their preparation.

> *"The culture of teaching and the organization of schools also serve as obstacles to effective field-based teacher preparation. Schools are not organized for teachers to work together on problems of practice in serious and sustained ways. With no tradition of inquiry, collaboration, or experimentation, there is a strong press to maintain the status quo."* (Feiman-Nemser, 2001, p. 1014).

Based on the above discussion, it has been established that foundational knowledge in educational linguistics and SLA should be in the epicenter of TPPs to produce informed teachers who can work effectively with ELs (Fillmore & Snow, 2000; Bunch, 2013; Nutta et al., 2012). Fillmore & Snow (2000) particularly stressed the need for all teachers to retain the basic knowledge of regular and irregular forms and their relationships, sociolinguistic variation in language use, historical linguistics to understand the complications in English spellings, knowledge of measures of English language proficiency (e.g., WIDA levels) and their characteristics, importance of interaction with non-ELs, and the knowledge and procedure of explicit teaching of English in Academic settings. Furthermore, new research has established the above requirements and has argued for some unique needs, e.g., linguistic responsiveness, focus on SLA principles, knowledge of the distinction between social and academic language proficiency (Lucas & Villegas, 2013; Bunch, 2013; Valdes, 2005), comprehensible input, classroom interaction for an authentic communicative purpose (Nutta et al., 2012; Lucas & Villegas, 2013), knowledge of the importance of transfer from L1 to L2; and effect of L2 anxiety and affective filters (Lucas et al., 2008).

However, before we reach a logical conclusion, it is vital to answer pressing questions related to proposed components of inclusion in TPPs. There are various reservations about the breadth and scope of topics that TPPs can practically embed in teacher preparation curriculum, how to prioritize the areas, and how basic linguistic knowledge can most effectively be taught to both pre- and in-service teachers (Bunch, 2013).

These doubts require serious consideration and in-depth studies, as noted and inquired by many researchers. First, the question of the time length it takes to provide an adequate introduction to basic language knowledge and wherein the course of teacher preparation this instruction should occur were examined from different perspectives, e.g., the quantity of educational, linguistic courses (Fillmore & Snow, 2000); space and place for such courses in already intensive teacher preparation curriculum (Baca, & Escamilla, 2002; Gollinck, 2002; Richardson, 2002); including alternative option, i.e., including language and literacy-focused

PST curriculum and continuous professional development opportunities throughout their in-service career (Baca & Escamilla, 2002; Gollnick, 2002; Valdes et al., 2005; Bunch, 2013). Furthermore, the questions about whether a teacher requires a formal or foundational linguistic knowledge such as linguistics or SLA have been addressed from different perspectives, e.g., the relationship between formal and foundational knowledge (Richardson, 2002); practical linguistic expertise needed to teach effectively (Feiman-Nemser, 2008; Hammerness, Darling-Hammond, & Bransford, 2005; Korthagen & Kessels, 1999); and appropriate linguistic knowledge-base for ELs and culturally diverse students in particular (e.g., Andrews, 1999, 2003; Freeman & Johnson, 1998, 2005; Trappes-Lomax & Ferguson, 2002; Tarone & Allwright, 2005; Ball, 2009; Faltis et al., 2010; Bunch, 2013).

As far as the issues of appropriate components of linguistics in the mainstream TPP curriculums are concerned, many theories have been developed and tested, e.g., concrete knowledge of socio-cognitive approach and strategies (Atkinson, 2011); understanding of the interaction between ELs' innate individual learning ability and environment (Lightbown & Spada, 2006); knowledge of the semiotics and ecology of language learning (van Lier, 2004); explicit teaching knowledge of grammar with the merging of conceptual and cognitive approaches (vanPatten & Williams, 2007); and knowledge of heritage language reacquisition/development and required linguistic knowledge around it (Valdes & Geoffrion-Vinci, 2011). Whereas, comparatively anachronistic and slightly more controversial theories, e.g., *monitor theory* postulated by Krashen (2003); the concept of basic interpersonal communication skills (BICS), and conversational, academic language proficiency (CALP) by

Cummins (1981, 2000) under *threshold hypothesis* have always guided TPPs in the United States and the length and breadth of linguistic components that PSTs need to master (Bunch, 2013).

Finally, in the era of Common Core and other standards, teachers face a challenge to support ELs in meeting content area expectations that call for rigorous use of language and literacy in English. A line of research dedicated to these concerns offers in-depth acumens on how all teachers should be primed for countering this challenge through TPPs and how PSTs can retain such knowledge throughout their vocations. Regardless of their content areas, to coach PSTs to engage ELs effectively with growing language and literacy prospects, TPPs' curriculum necessitates the development of pedagogical language knowledge, i.e., as Bunch (2013) puts it,

> *"Not to teach English in the way that most mainstream teachers may initially conceive of (and resist) the notion, but rather to purposefully enact opportunities for the development of language and literacy in and through teaching the core curricular content, understandings, and activities that teachers are responsible for (and, hopefully, excited about) teaching in the first place." (Bunch, 2013, p. 298)*

This idea is further supported by the studies conducted in content area courses and STEM TPPs. Parallel with the findings of SLA best practices for ELs, such studies reported that traditional STEM TPPs were detrimental to ELs because they ignored the interconnectedness of L1, L2 and cultures; belief that language learning precedes content learning; the faulty notion that language learning could be meditated through vocabulary and grammar drills; and the idea that language of mathematics and science were universal (Francis & Stephens, 2018). To this end, we can conclude that robust content can be accessible to all students through teaching strategies by providing multiple pathways to the understanding of language and content

(Gonzalez & Darling-Hammond, 2000). Most recently, Francis & Stephens (2018) noted a list of seven evidence-based points of consensus—relevant to supporting to all teachers of STEM subjects who work with ELs irrespective of their discipline, grade-level band, the language proficiency of their students, or their level of experience—in a Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine chronicle (in print) that include,

(a) explicit integration of STEM content and disciplinary language

(b) use and adaptation of Reform-Based curriculum

(c) shared professional multilingual instructional approaches in STEM classrooms

(d) engagement with families

(e) use of culturally sustaining pedagogies and explicit attention to equity

(f) targeted teacher learning around common societal biases and beliefs (Francis & Stephens, 2018, p. 6-10)

Furthermore, we know several things about the demographic shifts and its distinct challenges in everyday classrooms: contemporary changing standards comprise language and literacy needs that are challenging for all students, but more so for ELs; such problems call for shifts from present-day instruction for ELs; preparing teachers to carry out these shifts must become a mainstream concern; an EL is a composition of a heterogeneous population; and one of the implications of all of the above is that all teachers need to know something about language. However, it is less clear that what aspect of language the mainstream teachers need to learn, and how TTPs best prepare teachers to develop this knowledge (Bunch, 2013).

Nevertheless, language is not yet considered to be of prime importance in their professional practice by most mainstream teachers. A notion that there is a need for an addition of pedagogical language knowledge for mainstream teachers in existing (technological) pedagogical content knowledge (Bunch, 2013; Nutta et al., 2012; Francis & Stephens, 2018) is slowly getting traction among TPPs. This requirement is felt based on the difficulties and inadequacies underlying (technological) pedagogical content knowledge, and its flimsy impact in mitigating the wide-ranging learning gap between student groups (Bunch, 2013). The so-called wisdom of practice as Shulman (1986) noted, highlighted the blending of content and pedagogy to decipher overall aspects of a topic and how it is presented to diverse learners with the use of proper curricula, print materials (e.g., textbook) and lately, the technology (Shulman, 1986). As the school demography is becoming more diverse and the special-needs and linguistic minority student population continue to rise, it is desirable that the generalist PSTs know something about linguistics, second language acquisition, bilingualism, and so on (Bunch, 2013). On the basis of this knowledge, it is easy to recognize the sentiment of what Nutta et al. (2012) noted: "All teachers are expected to contribute to EL's academic achievement and language development…with its emphasis on language development expanded to the application of language skills in content areas, now resembles a more individualized and supportive mainstream class that is targeted to the specific needs of ELs" (Nutta et al., 2012, p. 9).

So far, this discussion has highlighted the academic achievement gap between students of different demographic backgrounds with a special focus on ELs and non-ELs. It also provided evidence of the significance of such differences in the present and future contexts and gathered information from multiple empirical studies that guide TPPs in preparing knowledgeable and capable PSTs to work with ELs. Since the last few decades, results of a vast number of studies provided strong evidence that language fluency is not a pre-requisite for content area study, thus, ELs should be allowed to learn academic subjects (e.g., science, mathematics, social studies, etc.) and language simultaneously, not separately or sequentially (Nutta et al., 2014; Francis & Stephens, 2018; Lee, 2018). During this period, these findings were supported by respective educational policies in the United States (E.g., the Civil Rights Acts of 1964, the Bilingual Education Act of 1968, the Equal Educational Act Opportunity Act of 1974, and the No Child Left Behind Act of 2002). Further response to this requirement, as noted by a plethora of research takes us to what we call now, an inclusive model of teacher education.

Though integration or infusion curriculum is new to TPPs, it has been a common practice in many content areas, e.g., instructional technology, exceptional education including mathematics, science, and social studies education as a stand-alone instruction of the skills. These strategies resulted in improved TPPs preparation; however, a bulk of research found that independent skill instruction for linguistically and culturally diverse learners may fail to achieve its goals because of the danger of sidelining the targeted skills. It prompts a need for more systematic and broad curricular development that contributes to all generalist PSTs' preparation to teach and assess ELs (Goldenberg, 2008; Nutta et al., 2012).

Definition, and Brief History of Infusion Model in the State of Florida and its Effectiveness

As the call for restructuring of TPPs to meet the guidelines and standards for both special and general education certification was widely recommended, a more prevalent initiative to improve the readiness of graduates for inclusive education inspired discussion, debate, and structural change in TPPs. With a need for modified TPPs to better prepare PSTs to successfully work with linguistic minority students at varying English proficiency levels, the Office of Bilingual Education and Multilingual Affairs at the U. S. Department of Education offered support to the universities that prepared teachers to meet the growing need for skilled and knowledgeable generalist teachers in teaching ELs. Institutions of higher education in the United States and Canada took quick actions, and they came up with a variety of models to enhance teacher practices.

Along the same line, the State of Florida mandated that the instructional and administrative personnel get specified training in line with the recommended best practices (25 ESOL competencies and skills which were later grouped under five broad categories) for ELs in the year 1990 by a Consent Decree between the State Board of Education and the League of United Latin American Citizens (LULAC) (Nutta et al., 2012; Wheeler & Govoni, 2014; Bristor, Pelaez, & Crawley, 2000). The milestone of the infusion model was carved by the state-mandated standards for the Recognition of Initial Teacher Preparation in P-12 teacher education under five ESOL domains. They were:

Domain 1: Cross-Cultural Communication and Understanding;

Domain 2: Language and Literacy (Applied Linguistics);

Domain 3: Methods of Teaching ELs;

Domain 4: ESOL curriculum and material development; and

Domain 5: ESOL testing and evaluation

(Nutta et al., 2012; Wheeler & Govoni, 2014)

It maintained that PSTs would be able to apply actively and demonstrate their acquired knowledge in these domains upon successful implementation of the EL-infusion model and clear elicitation of ESOL standards by teacher educators in higher education. It also highlighted the need of various field placements for the mastery of learned knowledge throughout teacher education (Van Laarhoven, 2007; Boyd et al., 2009; Smith, 2011; Nutta et al., 2012; Conderman et al., 2012; Gehrke & Cocchiarella, 2013; Wheeler & Govoni, 2014).

To borrow the words from Laarhoven et al., (2007), inclusive education entails the widespread practice of including all students (e.g., students with exceptionalities) in general classrooms. To define the phrase English for the Speakers of Other Languages (ESOL) Infusion in the context of teacher education, this study found Nutta et al. (2012) to be less ambiguous and complete. For them, ESOL Infusion is 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.” They further explain that EL-contents are comprised of “EL-focused topics, objectives, instructional materials and media, in-class activities, course assignments, field/clinical experiences, and assessments” which can be built into ‘EL-specific courses’ and ‘distributed across other teacher preparation courses’ (Nutta et al., 2012, p. 20).

As this trend gained momentum, North American universities, including *Boston College, University at Albany, Ontario Institute for Studies in Education, Miami University of Ohio*, were a few higher education institutes that directed their TPPs to infuse EL issues and teaching strategies. In the state of Florida, Florida Atlantic University (FAU) was first to embark on this journey when it received the full approval for an elementary education/ESOL endorsement degree program in 1990. Under the project called ESOL Integration Curricular Model (EIMC), FAU designed the first infusion curriculum to prepare PSTs in elementary education programs (Govoni, 2011; Wheeler & Govoni, 2014; Nutta et al., 2012). Many universities followed by designing ESOL infused core curricula by blending distinct programs for special and general educators into a unique TPPs in which all PSTs "undertake an expanded program designed to meet the guidelines and standards for both special and general education certifications" (Laarhoven et al., 2007, p. 440). As the momentum grew, the Florida Department of Education amended its minimal requirement for all PSTs to meet the following regulations:

1. Fifteen semester hours of ESOL endorsement for all PSTs majoring English Language Arts, Exceptional Education, Early Childhood, Elementary education;
2. A three semester hours course that addressed at an awareness level the five ESOL specified areas listed in the Consent Decree, (Govoni, 2011; Nutta et al., 2012)

Also, the State Board of Education voted the 25 ESOL Performance Standards into the law in the year 2009, which maintained that TPPs require instruction of ESOL courses and field/experiences, including observation, practice, and competency-based demonstration in teaching strategies for the teaching of ELs. With such modification in TPPs, ESOL experts in the national and state level worked tirelessly for two years to develop a new set of standards for school districts and TPPs (Govoni, 2011; Nutta et al., 2012).

ESOL Infusion One Plus Model of Teacher Education

The inception of the ESOL Infusion One Plus model was initiated by a contract of Nutta, Joyce, with the Florida Department of Education in the year 2001 to coauthor a manual that laid out a set of policies and criteria to help TPPs in submitting proposals for approval of infused ESOL endorsement programs. However, the formal initiation of ESOL Infusion One-Plus model lies beneath the widespread and thorough studies of pertinent issues under the project ESOL Training for all Preservice Educators Stressing Technology-based Resources (TAPESTRY), a five-year Office of English language Acquisition (OELA) funded initiative which "aimed to create a variety of materials and resources for faculty development as well as provide other support for embedding EL content into teacher preparation courses" (Nutta et al., 2012, p. 35). However, the scope of the project latter encompassed overall aspects of developing and implementing the infused ESOL endorsement program at the University of South Florida. After a rigorous multiyear effort of the whole team, followed by multiple action research studies using the data on the various fronts, e.g., faculty surveys, student assignment and portfolios, individual interviews, and focus group sessions, the team pinpointed the widespread issues of infusion model in all areas that included, embedding EL content in general teacher education courses, faculty development, candidate assessment, clinical and field experiences with ELs, and other program-specific challenges (Nutta et al., 2012). Once the plethora of resources was developed and the infusion procedures were carefully established, the team surveyed and consulted similar program coordinators and faculty within and outside the state of Florida and the whole host of actions culminated in the germination of ESOL Infusion One Plus Model.

Inspired by three vital communicative elements of interactionist approaches to SLA (a) receive comprehensible input, (b) affordances to produce meaningful output, and (c) participation in interaction that benefits and accelerates language learning, for ELs' second language development, a collaborative instructional model for ELs, the ESOL Infusion One-Plus model, was designed to meet PSTs' specialized expertise and preparation to enhance ELs' academic achievement. This model is flexible because it allows TPPs to assemble EL strategies, courses, and practices as needed. It is comprehensive because of its approach in incorporating all critical elements of teacher preparation such as courses, "field/clinical experiences, assessment, faculty development and scholarship, program administration, evaluation, and accreditation" (Nutta et al., 2012, p. 11).

Unlike the Florida Infusion model, the One-Plus model does not only focus on the preparation of the ESL specialist (i.e., those who learn the ESOL endorsement) but the training of generalist teachers of all subject areas and grade levels. Promoting three critical qualities of interconnectedness, cohesion, and interdisciplinarity, it endorses comprehensive, curriculum-wide infusion elevating participation of ESOL faculty and specialists with content area faculty, administrators, and other stakeholders (Nutta et al., 2012).

Though the general Florida model inspires the One-Plus model, they differ in some crucial ways. First, it is more flexible in terms of course requirements based on PSTs' area of study. The framework allows to achieve an appropriate level of depth for all programs that lead to EL-qualified credentials (EL-qualified for an academic subjects), with additional requirements for those who teach language arts. In addition, the PSTs who plan to specialize in ESOL and earn the ESOL endorsement or certification more course work is recommended. It also includes a 'Basic Coverage' option for the TPPs that are in the states where ESOL endorsement is not a requirement, and the program is unable to add coursework to their existing curricula. This option, 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).

Second, the Florida model does not specify the timing of field & clinical experiences but requires the PSTs to conduct teaching to one or more ELs under the supervision of ESL-endorsed teachers, however, the One Plus model recommends that PSTs complete a varying level of early field experiences in EL-embedded and EL-specific courses plus a clinical experience with ELs based on their major. Third, the Florida model uses the Florida ESOL standards to measure the competency of its PSTs, while One plus model uses EL Curricular Competencies based on the National Clearinghouse for English Language Acquisition (NCELA) roundtable recommendations (Nutta et al., 2012, p. 42).

The theory of interdisciplinarity guides One Plus model's conceptual foundation on three strong theoretical leanings, (a) curriculum development process, (b) the content to embed in academic subjects is informed by the latest instructional theory of 'effective instruction and assessment of ELs', and (c) its operation by the proven theories of 'instructional design' (Nutta et al., 2012). Interdisciplinarity refers to the collaboration of teacher educators from different disciplines and working together to enhance the chances of ownership after infusion. The theory of effective instruction and assessments of ELs maintains that 'teachers of ELs require specialized knowledge and skills in teaching and learning to respond to ELs' unique linguistic and cultural needs and the changing classroom contexts' (Coady et al., 2015, p. 344). And, using the tool of instructional design, the One Plus model aims to propel the effectiveness of infusion and to direct the incremental additive process higher than that of the broad conceptual model of curriculum design (Nutta et al., 2012). Thus, as noted by Nutta et al. (2012), the One-Plus model is an extension and refinement of the Florida Infusion approach.

The unique nature of the One-Plus model is that it does not only focus on the preparation of the English as a second language specialist but the training of mainstream teachers of all subject areas and grade levels as well. It is clear that One-Plus is a process-oriented model (e.g., faculty development, EL issues infusion, candidate assessment, clinical and field experiences with ELs, and other program-specific challenges) which requires at two three years to be implemented in any TPPs.

As most of the infusion modalities are still in the testing phase, we lack a strong research basis for understanding which model meets the challenges of changing needs (Boyd et al., 2009; Nutta et al., 2015). However, the preliminary findings of such programs, especially from individual case studies and small group qualitative and quantitative studies, seem to outperform the benefits of other TPPs (Smith, 2011; Conderman et al., 2012; Allday et al., 2013). This review compiled relevant journal articles and empirical studies that investigated the infusion models holistically, and partially, and have enlisted their findings to assess its effectiveness or any issues that they raise.

Some of the researchers (Gandara, Maxwell-Jolly, & Driscoll, 2005; Boyd et al., 2009; Smith, 2011; Allday, Neilsen-Gatti, & Hudson, 2013; Gehrke & Cocchiarella, 2013; Lavery et al., 2019) completed comparatively large-scale studies involving Infusion PSTs, infusion graduates, TPPs that embedded EL issues, or in-service teachers with(out) knowledge of EL issues. After conducting a multiple Analysis of Variance (ANOVA) 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 preparation had statistically significantly lower perception of their knowledge and skills, $\Lambda = .39$, $F_{(2,257)} = 192.99$, $p =< .0001$, $< \alpha = .025$ during the pre-course measures. The difference in perception remained comparable during the post-course measures. However, the researcher did not find any significant differences in PSTs' positive attitude towards EL inclusion based on the amount of ESOL courses they took in both pre-course, $F_{(1, 258)} = .01$, $p = .9279$, and post-course, $F_{(1, 212)} = .011$, $p = .7387$, measures. She concluded that more confidence in infusion PSTs' knowledge and skills to work with ELs was tied to the total time they were in the program.

A similar study was conducted by Gherke & Cocchiarella (2013), which surveyed 125 PSTs enrolled in Secondary Education (SE), Special Education/Dual Certification (SED), and Elementary Education (EE) majors. They were entitled to a varying degree of EL focused courses, field experiences, and student teaching requirements. Elementary Education has comparatively lower requirements (that is, two semesters of 74 hrs. each field experiences and two special education-focused coursework), while Special Education/Dual Certification required two semesters of full-time teaching, including two semesters of field experiences before their teaching requirements. On the other hand, the Secondary Education track required PSTs to complete three semesters of 74 hours each field experience, a full semester of student teaching plus two special education courses. The results showed that the PSTs' ability to identify characteristics of efficient inclusion was directly connected to the components of TPPs, i.e., PSTs from SE who had to go through the most extensive requirements were found to be more capable of identifying such elements and had higher degrees of positive attitudes toward the program curriculum.

Collecting data from 31 New York TPPs, approximately 36,000 students and their teachers, and the numerous TPPs faculties, Boyd et al., (2009) discovered that (a) on average, TPPs that produce teachers who were more effective at increasing student learning produced PSTs who were more effective in ELs; (b) the TPPs that gave higher importance to the portfolio of work done in the classroom with students, and field experiences had statistically significantly higher student learning outcomes than the ones that did not; (c) content-specific coursework in mathematics was more positively associated with teachers' value-added in the 2nd year compared to their 1st year of teaching. Boyd and the team concluded that TPPs that focused the classroom work during PSTs' first year as teachers and provided opportunities to for PSTs to study what they would be doing as 1st-year teachers, have prospects during their preparation to engage in actual practices involved in teaching, and who have a say on curriculum review performed better in terms of their students' learning gains in both math and ELA.

Similarly, Gandara et al. (2005) surveyed 5,300 California teachers to identify their challenges, experiences, and professional development need in the context of growing diverse academic, language, and social needs of ELs. They reported that the teachers with the highest levels of preparation were statistically significantly more likely to put together extra materials and attention required to ensure that ELs learn both content and language. They were found to communicate more with ELs in a relaxed manner, and thus, they had higher chances of knowing whether their students were learning. Those teachers took more opportunities to talk to the parents of ELs and help their colleagues compared to the ones who had little to no training and knowledge of EL issues.

One of the previous works that directly measured the One Plus model PSTs' effectiveness in fostering ELs' achievement comes from the teamwork of Lavery et al. (2019). The researchers compared pretest and posttest scores of 8,326 (n = 8, 236) students enrolled in the classrooms of 288 PSTs (n = 288) whose preparation implemented the One-Plus model to identify if they received enough training to support EL achievement in mainstream K-12 classrooms. They classified the PSTs in two groups that receive a varying degree of ESOL preparation contingent upon their majors. The PSTs who were pursuing content area majors (e.g., secondary math, science, and social studies) took six preparation courses that embedded 15 EL specific assignments, while the language arts PSTs completed twelve to fifteen courses that embedded 41-50 key ESOL related tasks (Lavery et al., 2019). Using a two-level covariate adjustment model, they attempted to decipher if EL unit post-assessment scores for One Plus content area or language arts PSTs were related after controlling for student preassessment scores, disability status, minority status, and *FRPL* status.

They reported that ELs demonstrated lower classroom preassessment scores (M = 36.1, SD = 23.2) than non-ELs (M = 41.7, SD = 24.7), $t_{(8324)} = -7.48$, $p < .001$, $d = -.23$; and lower postassessment scores (M = 69.7, SD = 21.6) than non-ELs (M = 75.4, SD = 20.5), $t_{(8324)} = -9.06$, $p < .001$, $d = -.28$. When the pre-assessment scores were included as covariates, the difference of marginal estimated means (at Pre = 40.8) between EL (M = 71.3, SD = 18.8) and non-ELs (M = 75.1, SD = 18.8) were statistically significant, $F_{(2,\,8323)} = 895.97$, $p < .001$. Student level variables significantly explained their post-assessment scores at the alpha level of a = .019. Among the student level variables, disability, $t_{(6401.1)} = -9.52$, $p < .001$, SES, $t_{(355.7)} = -5.65$, $p < .001$and EL status, $t_{(7690.7)} = -3.25$, $p < .01$, were among the significantly higher predictors of students' post-assessment scores. The difference between content area teachers and language arts teachers were statistically non-significant, $t_{(8290.2)} = 1.65$, $p = .098$, (Lavery et al., 2019, p. 10-12).

Finally, the researchers concluded that the ESOL Infusion One-Plus model could incorporated adequate curricular, instructional, and assessment skills could foster comparable pre-/post-assessment gains between EL and non-ELs. Furthermore, they highlighted that One-Plus mode would be able to mirror the PSTs' experience by depicting EL education and assessment in mainstream classrooms by infused programmatic courses and yet, providing more profound development in EL-specific courses. Overall, this model meets the requirements put forth by prior studies including Darling-Hammond (2008) that requires to house best practices for creating a powerful program (a) integration of theory and practice, (b) faculty collaboration, (c) courses incorporated assignments and performance assessments, and (d) a student-teaching placement (Darling-Hammond, 2008).

In addition, A small-scale two-level hierarchical linear model was conducted using the TWS data from Fall 2016 of One-Plus PSTs as a pilot study before the inception of this study. Total of 5,469 (n = 5,469) were nested within 236 (N = 236) One-Plus PSTs, with an average of 18.24 students per PST. There were six student-level predictors (a) student – the total number of students per class, (b) grade- grade-level of students as reported by the PSTs, (c) gender, (d) Ethnicity- students ethnicity was recognized as either White, Hispanics, Black, Alaskan, or Mixed race, (e) Free or reduced-price lunch eligible student as a proxy to students' SES, and (f) EL status. There were only two PST level variables (a) Program of study- a total of 3 programs of study were identified, i.e., elementary education, math education, and English language arts, and (b) PSTs grade level of teaching. Two different outcomes were modeled (a) pretest model to calculate the students' existing knowledge of the subject matter, and (b) gain model to assess the average learning gain after the treatment. Total gain per student was drawn from the difference in cumulative pretest scores and posttest scores. The descriptive analyses found that the average pretest score was 45.18 (M = 45.18) with a Standard Deviation of 24.16 (SD = 24.16), average posttest score (M = 82.48, SD = 17.13), and the total gain per student to be (M = 82.48, SD = 17.13) (Ghimire, Purmensky, & Regmi, 2020).

The analysis was conducted in three phases starting from the unconditional null model. This model provided an important statistic known as the intra-class correlation, which was used to measure the variation of results in student and PST level factors. Approximately 48% of pretest scores and 43% of posttest scores were found to be accounted for by PST level variables. Comparison between the ICCs of random intercept fixed slope models for pretest scores revealed that controlling all Level 1 variables would result in a 1% increment in teacher-level variance. The final model was identified after an exhaustive model building exercise. The final structure model could be expressed by following Level 1 (Student level) and Level 2 (PST level) variables. The gain model was simply the replication of the best fitting pretest model (Ghimire et al., 2020).

Level-1 Model

$$PREPERCE_{ij} = \beta_{0j} + \beta_{1j}*(STUDENT_{ij}) + \beta_{2j}*(GENDER_{ij}) + \beta_{3j}*(FRLUNCH_{ij}) + \beta_{4j}*(ELL_{ij}) + r_{ij} \quad (1)$$

Level-2 Model

$\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

$\beta_{2j} = \gamma_{20} + \gamma_{21}*(PROGRAM_j) + \gamma_{22}*(GRD\_LVL_j)$

$\beta_{3j} = \gamma_{30} + u_{3j}$

$\beta_{4j} = \gamma_{40}$ (2)

As seen in the final model, the student level predictors (e.g., the total number of students per class, students' grade level, and ethnicity) that were not significant in predicting pretest scores were not included in the study. Teacher level variables did not interact statistically significantly with any of student-level predictors but gender; thus, they were not included with other predictors.

The results suggested that ELs had statistically significantly lower classroom pretest scores, $t_{(4755)} = -6.06$, $p < .001$ compared to non-ELs. Similarly, FRPL eligible students, $t_{(4755)} = -5.37$, $p < .001$. Conversely, smaller class sizes had statistically significantly higher pretest scores, $t_{(235)} = 8.12$, $p < .001$, and male students had comparatively higher pretest scores, $t_{(4775)} = 2.48$, $p < .05$. Furthermore, students' gender interacted significantly negatively with teachers' grade level of teaching $t_{(4775)} = 2.48$, $p < .05$. In other words, the higher grade the PSTs taught, the higher the chances of students' gender interact with their learning outcomes.

The final gain model shed light on students' learning gain after treatment. The average gain per student, $t_{(235)} = 34.17$, $p < .001$, was statistically significantly higher than zero. All students had higher linear posttest scores; however, some gains were much higher than others. The differences in test scores based on gender still existed, $t_{(4990)} = .41$, $p > .05$, but it was not statistically significant. The bigger class size was still associated with a lower test score, $t_{(235)} = -.431$, $p < .001$, and it was statistically significantly different from zero. On the other hand, the drastic results could be seen in the learning gains of ELs and FRPL eligible students. The learning gains were much higher in FRPL eligible students, $t_{(235)} = 3$, $p < .01$, than non-FRPL students. Similarly, ELs' had significantly higher learning gains, which was comparable to that of non-EL students. There still existed a small negative difference, $t_{(4990)} = -.25$, $p > .05$, but it was not statistically significant. Finally, the variance component statistics were statistically significant, $\chi^2_{(235)} = 4972.88$ and 704.71, $p < .001$, which suggested that the study was still missing some important Level 1 variables (Ghimire et al., 2020).

The findings of this pilot study were coherent with that of Lavery et al. (2019) and Smith (2011), in terms of the results of the program of study of PSTs (e.g., English language arts, science education, elementary education, etc.). All these studies found ELs, low-SES students to have statistically significantly lower achievement level on their pretests, and comparable results on posttest measures.

Similarly, Allday and the team evaluated a total of 109 TPPs that offer initial certification in elementary education from all over the United States based on their characteristics of disabilities as it related to special education, differentiation of instruction/inclusive practices, classroom and behavior management, and collaboration. These requirements are consistent with that of the ESOL One-Plus model. The team found that on average most TPPs required PSTs to take 124.39 (*SD* = 8.47) credit hours to graduate from the elementary education program, of which 59.80 (*SD* = 13.31) credit hours were dedicated to initial certification, and 2.35 (*SD* = 1.27) were assigned for special education. Likewise, an average of 1.12 (*SD* = 1.99) credit hours of training was directly related to differentiation of instruction, but an overwhelming percentage (i.e., 93%) of TPPS did not offer any course-specific to teacher collaboration. On the other hand, 27% fo the reviewed universities required at least three credit hours of studies that explicitly tied to special need students; 41% of the TPPs required at least 3-credit hours of classroom behavior management courses (Allday et al., 2013). Though these results painted a fragile line of hope, they are more positive results if we compare them with the similar studies in the past, e.g., Powers (1992), Harvey et al., (2010), etc.

There have been few small-scale studies that measured the effectiveness of either the Infusion model TPPs or some aspects of effective practices recognized by the infusion model. Teacher candidates who attend TPPs with higher requirements for special education (e.g., infused courses, course-specific competencies, clinical experiences) are associated with most substantial gains compared to the PSTs with a lower level of requirements (Laarhoven et al., 2007; Conderman et al., 2012). Furthermore, there are few journal articles that report teacher educators' personal experiences of attending professional learning, the process of embedding EL content into their courses and teaching such courses (e.g., Niday, 2011; Hutchinson, 2011; Isaac & Quantz, 2011; de Olivera & Athanases, 2007; Costa et al., 2005; and Meskill, 2005). Such papers provided essential recommendations for the faculties who want to infuse EL content into their areas of teaching.

CHAPTER THREE: METHOD

Introduction

Chapter Two reviewed the available literature on teacher education and how it holds the potential for narrowing the achievement gap between ELs and Non-ELs in the United States K-12 education system. This chapter discusses the research questions that guide this study and explains the research design, sample size, and data collection procedures and a hypothesized model to answer those questions successfully.

This study was carried out using a quantitative methodology with a non-experimental comparative design. The archival data were analyzed using a three-level hierarchical linear (HLM) modeling, two-level longitudinal HLM, and a MANCOVA procedure. This regression-based correlational quantitative research implemented HLM to quantify the relationship between students' sociodemographic characteristics and their pre- and posttest scores and how the PST level characteristics moderated students' performance. Based on the supporting literature and empirical studies, this study predicted that there existed a gap of prior knowledge (achievement gap) of a topic based on their sociodemographic status given that the teachers are not adequately trained to work with all students. The gap in pretest scores shed light on this issue. Assuming that the One-Plus PSTs possessed the *knowledge base* for equity education, *tools* (e.g., language knowledge, ESOL methodology, and assessment) to mitigate unequal practices, and *experience* to carefully implement the available resources, through service-learning, micro-teaching, clinical/lab experiments, and updated TPPs curriculum, this study predicted that the One-Plus

PSTs would help boost learning gains in all students and more so to low-achieving students including ELs thereby lowering the achievement gap.

This study involved collecting data on multiple variables for the participants and computing the extent of the simultaneous relationship among those variables at varying hierarchical and across levels (Woltman, Feldstain, MacKay, & Rocchi, 2012). It is not a unique phenomenon to have variable group sizes in social science research, which poses a threat to many statistical procedures in terms of the validation and generalization of the findings. Many statistical analyses (t-test, ANOVA, Chi-Square tests, etc.) cannot provide results with confidence if the sample size of one of the groups is much smaller than the other, but HLM can, even in the extreme condition in which the sample size (n_j) is not larger than 1 (Hahs-Vaughn, 2016). This design is highly useful in conducting inquiries related to education and social sciences because several variables in these institutions often influence a set of behaviors (Gall et al., 2007). In other words, this mixed-effect statistical model analyzed the variance of multilevel factors where some coefficients were fixed and others random to measure the effectiveness of PSTs' in increasing test scores and if the gains were strong enough to narrow the achievement gap (Snijders & Bosker, 2012).

This study further presumed that the narrower achievement gap between student groups would be consistent among various content area courses, in different classroom settings, among the diverse student population, by the PSTs' of different majors over time and it would provide enough confidence to credit the Infused ESOL One-Plus teacher preparation model. In addition, the results of the long-term time component (five-semesters) allow this study to compute the rate of retention of observed effectiveness over time.

To put it succinctly, this study tried to explore all possible aspects of the Infused ESOL One-Plus PSTs' effectiveness and their variability based on major, subject of teaching, and grade-level of teaching as represented by '*Z*' on students' learning gains '*y*' (fixed-effect model) in the leftmost figure below; effectiveness after controlling for student-level variable '*x*' (e.g., gender, FRPL-status, etc.) as shown in the middle, and their interactions as in the rightmost figure.

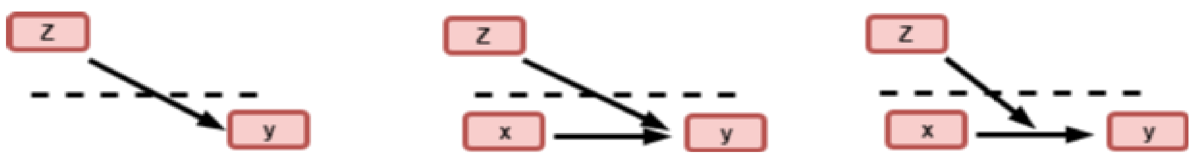


Figure 1. The proposition of effect of teacher-level variables 'Z' on students' test scores 'y' under various conditions.

Note. Adapted from "Multilevel analysis: An introduction to basic and advanced multilevel modeling' by T. A. B. Snijders, and R. J. Bosker, 2012, SAGE Publication, p.12.

Thus, the non-experimental correlational design was deemed an appropriate model for this study because of its capability to identify the relationship, including *within*, *between*, and *cross-case* impact of multi-level variables to answer one overarching research question and its two sub-questions on the effectiveness of Infused ESOL One-Plus teacher preparation program in terms of producing PSTs who are capable of narrowing the K-12 achievement gaps among student groups:

1. What is the rate of change of One-Plus PSTs' effectiveness in reducing the EL achievement gap over time?

    1.1. What proportion of the variance in students' academic achievement is associated with students' EL-status, FRPL-status, ethnicity, disability status, and gender?

    1.2. What proportion of the variance in students' academic achievement is associated with PSTs' major, grade-level of teaching, class size and the subject they teach?

The achievement gap between EL and non-EL K-12 students was of direct interest in this study. The ultimate purpose of this study was to measure PSTs' effectiveness in bridging the gap between EL and non-EL students. However, as seen in the research questions, variables like students' ethnicity, disability status, socioeconomic status, and gender were also included because they are the mandatory components of the TWS instrument used in the institution studied. Furthermore, they were used in this study to see how they interacted with students' EL/non-EL status and to identify their positive or negative contribution to the existing achievement gap.

Research Design

This quantitative non-experimental study proposed to use a correlational research design, and the data were analyzed using HLM. A separate comparative procedure was conducted using MANCOVA to measure the change in ELs' pre- and posttest scores as a function of other student and teacher level variables. For a long time, quantitative research was used with the belief that our world is relatively stable and uniform, and a systematic inquiry of numerical data allows researchers to describe, explain, predict, or control variables and phenomena of interest (Gay, Mills, & Airasian, 2009). This well-established research process focuses on objectivity to helps researchers to generalize findings outside the set of circumstances encompassed in the study (Fraenkel et al., 2012). A specific type of quantitative research—correlational research—investigates the relationship between two or more variables, recognizing trends and patterns in researched data (McMillan, 2008). In other words, this type of research is like detecting variables that seem to interact with each other so that a shift in one variable would allow us to measure the change in other variables.

In most cases, data collected in organizations tend to be nested due to the inherently hierarchical nature of organizations (Hofmann, 1997; Sinjders & Bosker, 2012), and the computation of such data is best analyzed using statistical techniques that account for hierarchy (Woltman et al., 2012). It is, thus, desirable to use multilevel analysis when the unit of statistical analysis involves different layers of individual and institutional functioning, for example, students in a classroom, are likely to be affected by their gender, age, and socioeconomic status including their classmates and teachers (Gall et al., 2007). As noted in the *Sage Encyclopedia of Social Science Research Method*, the nested design is a research design in which levels of one factor are hierarchically nested within levels of other factors (Lewis-Beck, Bryman, & Liao, 2004). The proposed statistical methodology for this study, HLM is a complex form of ordinary least squares (OLS) regression, which studies variance in the outcome variables when the predictor variables are at different hierarchical levels.

Nested research design offers a conceptual and analytical apparatus for examining and extracting conclusions concerning the impact of phenomena at different levels of analysis (Hofmann, 1997). Being open to including statistically significant random effect in the final model, this study attempted to account for the factors that were not directly listed either as Level 1 predictors or as Level 2 moderators, as well.

To put it succinctly, Figure 2 depicts the overall concept map for the proposed research design, which gives the wholistic idea of all variables and respective covariates in this project. As it shows, students were in the first level (Level 1) of the spectrum. They were the most important components in this study because the overall calculation was conducted on the ground of their pre- and posttest scores. Students were nested under the One-Plus PSTs, who were further nested under the semester of their internships.

This study took students' EL status, FRPL status, disability status, ethnicity, and gender in account and measured the differences in students' pre- and posttest scores based on such characteristics. Pretest scores allowed establishment of a baseline performance for the students. The variation in students' pretest and posttest scores shed light on the learning gains, i.e., the rate of change on students' test scores. Thus, this study accounted for the average rate of change in students' posttest scores from pretest scores, as well. The learning gains of all students of a One-Plus PST collectively resulted in his/her average teaching effectiveness. This study then tried to account for the differential effectiveness based on given moderators, i.e., the total number of students, grade-level of teaching, and major (. Furthermore, the average effectiveness of all teachers provided average semester wise effectiveness. Calculating average effectiveness per semester, this study tried to foreground the trend over time.

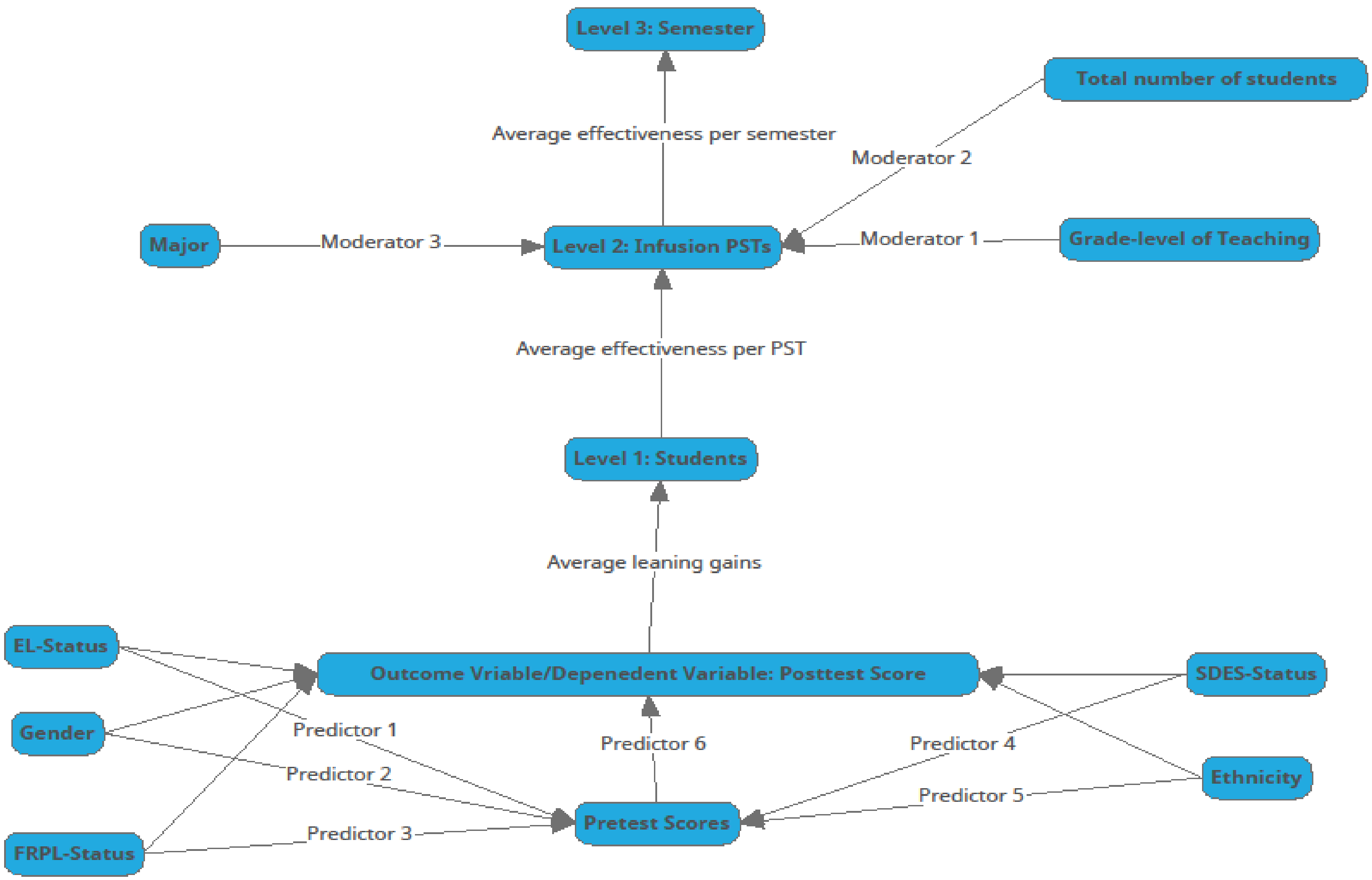


Figure 2. Proposed Research Design, Concept Map.

In other words, this study analyzed how the outcome variables (posttest scores, pretest scores) depended on the relevant (both student and PST level) covariates. As mentioned earlier, the major statistical tool was a 3-level cluster-randomized trial (3-level CRT) design randomized in both Level 1 and Level 2 because of the presence of the covariates in both levels. The three-level model consisted of three submodels, i.e., in this study, the Level 1 model represented the influence of student-level covariates, the Level 2 model captured the effect of teacher-level factors, and the Level 3 model presented the average effect of all PSTs who interned in a specific semester, thereby allowing the researcher to analyze the Infusion PSTs' effectiveness on the achievement gap, as shown in Figure 3.

Statistically, there were $i = 1, \ldots, n_{jk}$ student-level characteristics (demographic data), nested within each of $j = 1, \ldots, j_k$ One-Plus PSTs, which in turn were nested under each of $k = 1, \ldots, K$ semesters. The covariates in this study (e.g., students' EL-status, FRPL status, disability status, ethnicity, gender, and teachers' major, class size, and grade-level of teaching), meet the requirement of having a strong linear association with the outcomes, and they had similar association within each treatment condition because all the PSTs under question attended the same TPP. Thus, this study met the requirement for a 3-level CRT and expected the covariates to enhance the chances of precision (Moerbeek, Van Breukelen, & Berger, 2008).

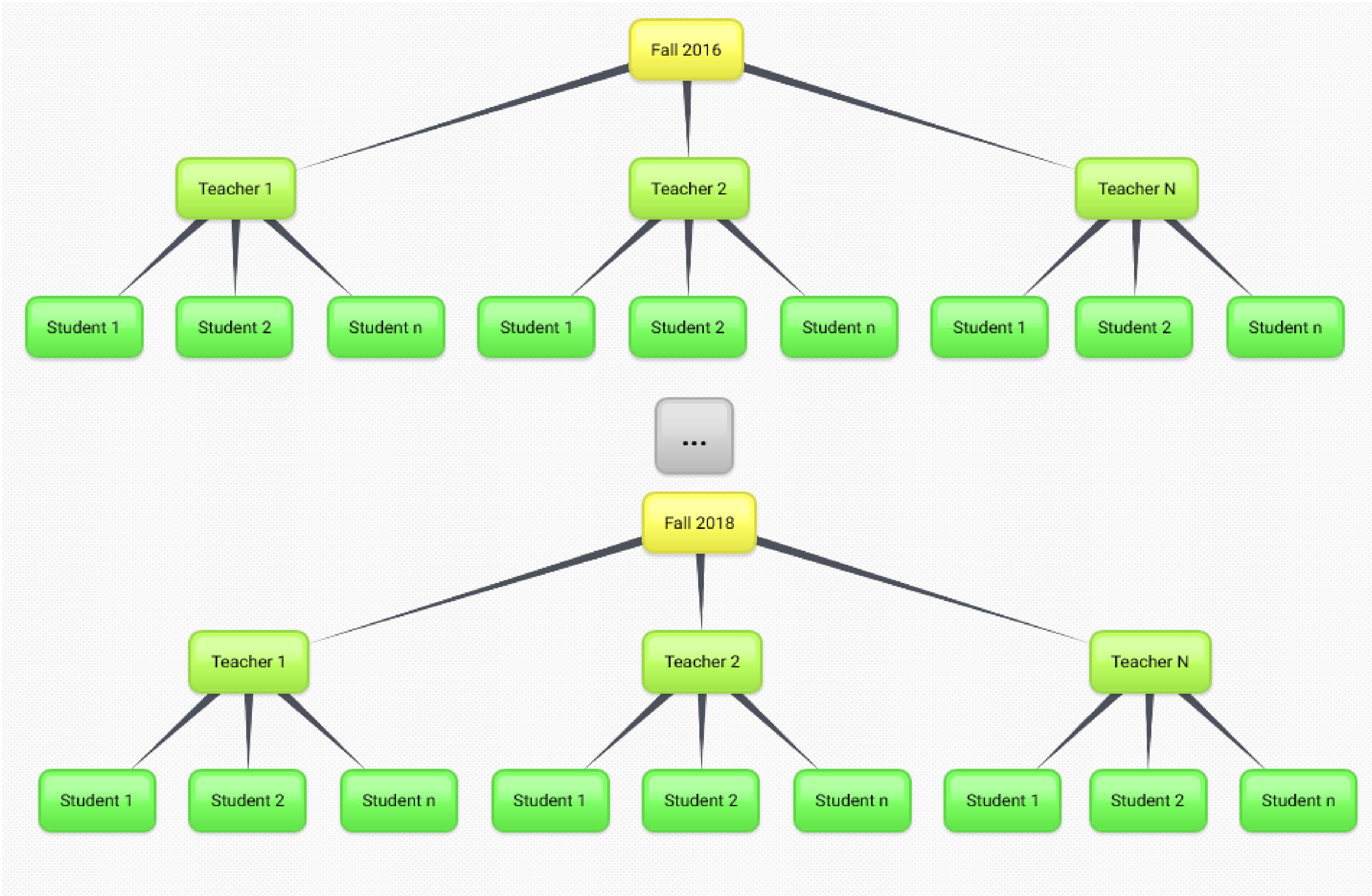


Figure 3. Three Level Hierarchical Design.

Data Sources and Instrument

This study utilized PST reported data collected through Teacher Work Sample (TWS), "a key assignment in all initial teacher preparation programs" (Lavery et al., 2019). Researchers acknowledge the TWS as a powerful instrument to provide evidence of teacher candidates' capability to design, implement, and assess standards-based student learning and to reflect on that learning process in an evocative manner (Foster et al., 2010; Denner et al., 2001; Denner, Salzman, & Bangert, 2001). In general, TWS is regarded as a professional culminating activity capable of integrating qualities like, (a) ability to locate and gather information, (b) critical thinking, reasoning, and analyzing, (c) effective communication skills, (d) ability to demonstrate human experience and understand them, (e) understanding cultural differences and the relationship, (f) integrate the breadth and diversity of knowledge and experience, and (g) ability to make informed, intelligent and value decisions (Benton et al., 2012).

The TWS instrument in the institution studied is composed of two separate sections: the prompt and rubric. The "*prompt*" section provides instruction and explanation on completing the different components of the TWS (e.g., contextual factors, learning goals, assessment plan, design for instruction, instructional decision making, analysis of student learning, and evaluation and reflection) (Fredman, 2004). The second part of the TWS is allocated to input the wide-ranging demonstration components by the PSTs. The TPP under study used TWS GraphMaker™ version 5.1.2 to collect students' demographic information, academic information (e.g., learning goals, etc.), and test scores (Version 5.1.2.; Lavery, 2012). In short, the TWS GraphMaker™ was a Microsoft Excel-based spreadsheet that stores and tracked the information of all required variables for this study. In general, the TWS is filled and approved digitally and submitted online by the PSTs to the institutional effectiveness division of the college. This study investigated and compared the pre-and posttest scores of various student groups taught by One-Plus PSTs during their student-teacher placements for five semesters starting Fall 2016 through Fall 2018.

To give some historical perspective about the TWS, it was initially developed at Western Oregon University to systematically and meaningfully capture student teachers' level of competency by connecting teaching and learning to be eligible for licensure (Schalock & Myton, 1988; Girod, 2002; Henning, DeBruin-Parecki, Hawbaker, Nielser, Joram, & Gabriele, 2010). In other words, TWS allows the PSTs to "to think about teaching in ways that are linked rightly and continuously to pupils' learning, to gain experience in teaching in this manner and to demonstrate effectiveness in doing so" (Girod, 2002, p. 20). At the institution studied, the PSTs are prompted to set at least three learning goals and appropriately align them to local, state, or national standards and justify them in terms of the required knowledge base. The contextual factor includes school or school district factors, and students' demographic information.

Furthermore, the PSTs create a series of assessment plans by aligning their pre- and post-assessments to their learning goals. They, then, design instructional plan as guided by their selected objectives followed by classroom instruction, and assessments. They are prompted to use the GraphMaker™ to input the pretest and posttest scores with a common structure and format, which generates various charts and graphs that include pre/post assessment scores (a) for all students, (b) by classes if PSTs teach more than a class, (c) by gender, (d) by ethnicity, (e) by socioeconomic status, (f) by disability status, (g) by ESOL status, and (h) by average class/group size (GraphMaker™ Version 5.1.2.; Lavery, 2012). The figure below shows a sample graph for pre/post-test scores by students' EL status.

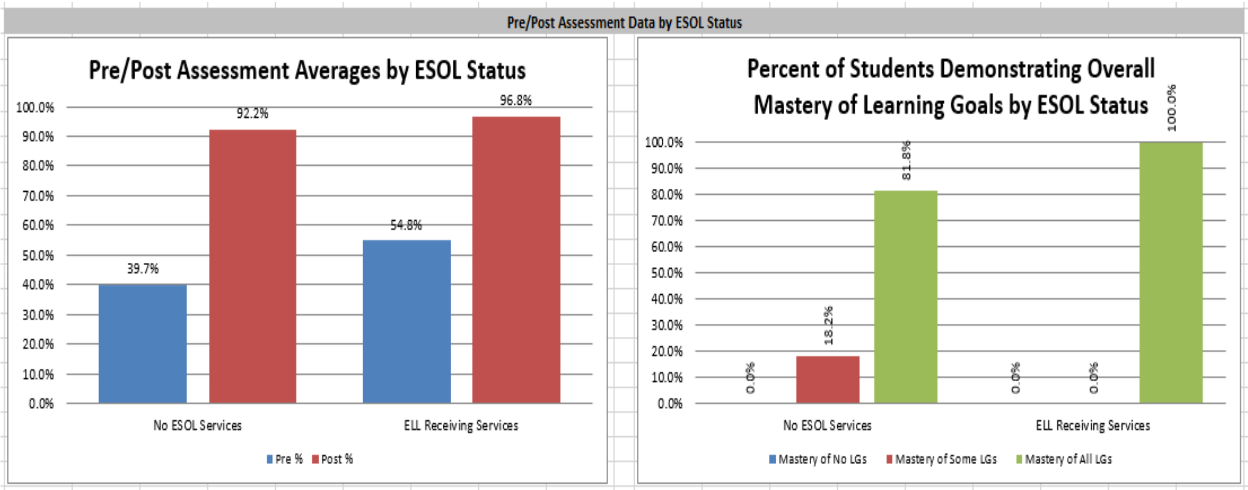


Figure 4. A Sample Graph Showing Pre/Post Assessment Data by ESOL Status.

Note. Adapted from Teacher Work Sample GraphMaker™, Version 5.1.2 by M. Lavery, 2012.

In general, the PSTs plan an instructional unit and design pre- and posttests for that unit based on the learning goals they came up with. They administer a pretest and then teach the unit. After completion of the unit, they administer the posttest to their students. The PSTs are trained to create pretest and posttest exams containing more than two items and resemble other assessments that the students are accustomed to taking. The pretest and posttest can be either (a) identical instruments, (b) identical subsets of items from similar instruments, or (c) similar performance tasks scored using identical rubrics (Lavery et al., 2019). The PSTs are educated that the same tests can cause testing effects, and they are encouraged to make changes on the posttests, as mentioned in the TWS manual. Thus, this study assumes that PSTs followed the test protocol, i.e., the tests "measure the same learning outcomes at the same level of difficulty using similar assessment items and/or methods and produce scores that can be compared to one another to show growth" outlined in TWS manual. Furthermore, to ensure that the changes/gains in the posttest scores be reasonably attributed to student learning and not to the memory of the pre-assessment, the PSTs make relatively subtle changes to their assessment.

The data set obtained from GraphMaker™ demarcated and stored student-level covariates, i.e., EL-status, free or reduced-price lunch status used as a proxy to socioeconomic status in this study, gender, grade level, disability status, and the content area of the PSTs in the distinct column under specific PST. The obtained scores were clearly listed under the pretest and posttest headings for each student that took the tests, and the cell were left blank if they missed any or all the tests. Furthermore, the TWS has dedicated columns to store teacher-level variables, i.e., major, subject of teaching, grade level of teaching, and the class size. The data set contained many other variables, PST ID, and some other units like registered internship section, and the school size, which were not included in this study.

A sample of archival data was assessed for a small pilot study before the initiation of this project. During the data screening process, this study found that the PSTs reported a range of pretest and posttest scores (minimum 1 and maximum 5). On average, the PSTs reported students' scores on three pretest and posttest each. The number of pretest and posttests were tied to the number of learning goals the PSTs identified for the content areas they taught. One set of pre-, and posttest data aligned to one learning goal, which means, three pre/posttest scores informed three distinct tests on three different units. Thus, it is desirable that the reported pretest scores per student be added and averaged together to create a composite pretest score, which were further changed into the percentage score; the same was done to reported posttest scores.

Sampling and Sample Size

There exist two types of population, i.e., target population and accessible population for any quantitative study (Gall et al., 2007). In this study, the target population was the One-plus PSTs and their students in a mainstream K-12 classroom. Thus, this study relied on purposive cluster sampling (Gall, Gall, & Borg, 2007) followed by random sampling. The study drew all available samples collected within a specific time frame, i.e., from Fall 2016 to Fall 2018 (5 semesters) and utilized all PSTs and their students if they met the study criteria. Thus, this study used the census, i.e., the entire population of PSTs who conducted their student-teacher practicum during the specified timeframe. A census is instrumental in providing data on all individuals in the population and eliminating sampling error (Israel, 1992). Most of the biases and standard errors in statistical analyses are due to the sampling variation (Jolly, 1969). With a sample of 100% or nearly 100% (because of missing and incomplete data), this study assumed to increase the statistical accuracy by eliminating the errors attributable to sampling. This study was not limited to a specific grade, school, class size, gender, subject, ethnicity, or major. It aimed to embed a wide range of attributes, qualities, and the situation in both teacher, student levels, thus it is the most appropriate sample for this study. In addition, this study used a cluster sample design because it strived to study all clusters in their entirety. Mathematically, the probability of selecting research participants for this study can be denoted by:

$$\pi_{i/j} = 1 \; for \; all \; i \; and \; j \qquad (3)$$

Where,

π_j = inclusion probability for students of a certain PST j

$\pi_{i/j}$ = inclusion probability for all One-Plus PSTs who conducted student-teacher practicum in one of the five semesters

Thus, the marginal probability of observing student-level unit *i*, in cluster *j*, is given by the product,

$$\pi_j \, \pi_{i/j} \tag{4}$$

Hierarchical Linear Modeling (HLM) was utilized to investigate the predictability of Level 1 variables on students' average pretest and posttest scores, their interactional impact, and whether the Level 2 variables moderated any of the outcome and the size of the effect. HLM concurrently explores relationships within and between tiered levels of grouped data making it more effective at accounting for variance among units at different levels than other existing analyses (Woltman et al., 2012).

Most of the empirical research in social and behavioral sciences establish and test the null hypothesis that the researchers hoped to reject to establish the facts about the phenomena under study (Cohen, 1992). This study, likewise, had propounded multiple null hypotheses and expects to reject them all to establish the facts expressed through research questions. This procedure is called power analysis in statistics. It is most useful if we conduct it while planning for a study because it helps researchers to investigate the relationship between the range of sample sizes that are considered possible, effect sizes assumed to be naturally important, level of variance in the population and anticipated levels of α and statistical power (Thomas & Krebs, 1997).

The question of an adequate sample size is difficult to answer even with simple single-level statistical analyses, and it gets more complicated in the case of multilevel models (Hahs-Vaughn, 2016). Most often, in HLM, the sample size of the highest level (i.e., semester level in this study) is of chief concern because the sample size at this level is always much smaller than other levels (teacher level, and student level). There are few recommendations available in methodology literature for minimum group sample size like Level 2 or Level 3 variables in this study. In most cases, the sample size for the lowest level, i.e., student-level participants, is of least concern and is recommended to retain. Some researchers recommend sample size be ten times larger than the number of formative indicators of a single construct (Barclay, Higgings, & Thompson, 1995; Hair et al., 2016; Hahs-Vaughn, 2016). Many researchers have criticized this proposition because it does not take critical components like model background and data characteristics in the account (Hair, Ringle, & Sarsted, 2011).

Identifying sample size at each level is not a separate procedure because it should always be considered in relation to the proportion of ICC and estimation methods like maximum likelihood (ML) or full maximum likelihood (FML). Thus, it is important to consider simulation research and the recommended variable sample sizes. Mass & Hox (2005) recommended the number of groups be (30, 50, 100) with size of each group (5, 30, 50), and ICC (.1, .2, & .3), respectively. Sinjders & Bosker (1999), on the other hand, tested the results of 10 groups and concluded that Level 2, with just ten groups, was insufficient. A recent simulated study was conducted for a sample of 60 and 200 teacher candidates by Lavery et al. (2019) and concluded that both sample sizes were sufficiently powerful to answer the research question they posed. Thus, on average, as Hahs-Vaughn (2016) mentions, there should be at least 20-40 Level 1 sample, and a minimum of 30 groups are "required to achieve the power of Level 1 predictors of .80 or greater" (Hahs-Vaughn, 2016, p. 522). With these analyses in mind, this study had a total of 768 Level 2 variables and approximately 27 students per PST, making the overall sample size of 768 (n_j = 768) PSTs, and 20,809 (n_i = 20,809) students. This sample size was much larger than any recommended sample size threshold.

To verify that this proposed sample sizes (number of semesters = 5, number of PSTs = 768, & number of students = 20,809), held necessary statistical power to answer the proposed research questions, a priori power analysis was conducted using Optimal Design (O/D) software. For this three-level cluster randomized trial with a continuous outcome variable, the given sample size should be enough to achieve the power of .80, with the effect size of at least .35.

Data Collection Procedures

This study used de-identified secondary (archival) data collected by a large public university in the Southeastern US. This data was not publicly available. As per the institution's dissertation protocol, an appropriate Institutional Review Board (IRB) approval was secured before accessing, collecting, and analyzing the data.

This study acknowledges the fact that missing data are common problems in social science research, and they have a huge impact on the results of statistical analysis if not dealt appropriately before analyses. Thus, after preparing data for the study, they were screened for missing cases, and appropriate steps were taken to deal with them. During the data cleaning/screening process this study encountered three different types of missingness. First, some of the GraphMaker™ files submitted were corrupt, they either did not open, or they were empty. Such cases were not included in the study. Second, in some of the cases the files the teachers submitted had no information on students' sociodemographic statuses (e.g., gender, EL-status, ethnicity, disability status, or socioeconomic status) and/or PST characteristics (major, subject, grade level, etc.). The cases that failed to report the key independent variables, including EL-status were cross-referenced with the variables of interest, and they were deleted case-wise. Third, in some cases, the data set had all required information but missed either pretest or posttest scores. In such cases, this study proposed to test the variable against 'missing completely at random' (MCAR) or 'missing at random' (MAR) assumptions and the missing data was substituted using the standard implementation method known as multiple imputations (Kontopantelis, White, Sperrin, & Buchan, 2017; Bhaskaran & Smeeth, 2014; Hahs-Vaughn, 2016).

The cases of a variable are missing at completely random (MCAR) when the missing cases are the random subset of all cases, and such the observed values share similar distribution (Bhaskaran & Smeeth, 2014). This phenomenon is considered a rare occurrence in social science research. Likewise, MAR accounts for the missing data, which may be systematically different from observed data, but they can be accounted for by other observed variables (Kontopantelis, White, Sperrin, & Buchan, 2017). Multiple imputations are considered the most appropriate methodology to account for the MCAR and MAR phenomena because it allows for with-in and between-estimation variance components by (a) extracting the missing data from their subsequent predictive distribution, (b) scrutinizing different dataset individually with a selected method including regression, and (c) merging the estimates and their standard errors across the *N* analyses (Kontopantelis, White, Sperrin, & Buchan, 2017). After careful consideration and cross checking, this study found less than 3% cases in which students' either pretest or posttest scores were missing. Those cases were replaced by the class mean.

Table 1.

Measures; Variable Descriptions.

Variable	Description
Level 1, Student-Level Variables:	
English Learner (EL) Status (ELS)	This is a categorical variable. The students enrolled in the ESOL program were marked 'Yes' and they were coded 1; the students who exited the ESOL program or completed the two-year follow up were put in 'EXIT' group and they were coded 2; and all other students were put under 'No' category, and they were coded 1.
Free or Reduced-Price Lunch Status (FRPL)	This is a dichotomous predictor. Students who were eligible for free or reduced-lunch program represented that low-SES students. They were coded 1, and the non-eligible students were regarded as the high-SES students. They were coded 0.
Students' Disability Status (ESE)	Students were classified either as students with exceptionalities, or gifted students, or as students without exceptionalities. The students without exceptionalities refers to the average students who did not have any psychological, emotional, or physical accommodation to help them learn in mainstream classrooms. Students with exceptionalities were coded 1; students without exceptionalities were coded 0; and the gifted students were coded 2.
Ethnicity (ETHNICITY)	This is a nominal predictor derived from the students' race, 0 for Whites, 1 for Black, 2 for Hispanics, 3 for Asian or Pacific Islanders, 4 for Alaskan Natives or American Indians, and 5 for Other or Multiracial students. For the multivariate analysis of variance (MANCOVA) this variable was dichotomized as a MINORITY variable. White students were marked as non-minority and all other students were put under minority groups. The minority group was coded 1, and non-minority was coded 0.
Gender (MALE)	This is a dichotomous Level 1 predictor providing information on the students' gender. Coded 1 for male students and 0 for female students.

Variable	Description
Pretest Scores (PRE_SCR)	A continuous percentage variable derived from students' performance on pretest/s given by the One-Plus PSTs before they taught a unit.
Posttest Scores (POST_SCR)	A continuous percentage variable derived from students' performance on posttest/s given by the One-Plus PSTs after they taught a unit.
Level 2, PST-Level Variables:	
Major (MAJOR)	This is a nominal variable. The PSTs who were admitted to the Elementary Education Program were marked ELEM and they were coded 0. PSTs in English Language Arts (ELAE), Math Education (MATH), Science Education (SCIE), Social Studies Education (SSED), Art Education (ARTE), and Foreign Language Development Education (FLDE) were coded 1, 2, 3, 4, 5, and 6, respectively.
Subject (SUBJECT)	This nominal variable refers to the subject or the content area the PSTs taught during their internship. Language Arts (LA), Math (MA), Science (SC), Social Studies (SS), Visual Arts (VA), and World Language (WL) were coded 0, 1, 2, 3, 4, 5, & 6, respectively.
Grade Level of Teaching (GRADE_T)	This is a categorical variable, which refers to the grades the PSTs taught during their internship. Grades K-6 were put together as Elementary grades (ELEM) and coded 0. The grades 7 and 8 were put as a separate category (MIDDLE) and coded 1. Finally, the grades 9, 10, 11, & 12 were put under high school (HIGH) category and coded 2.
Class Size (CL_SIZE)	This is a categorical variable that refers to the number of students per class the PSTs taught during their internship. Following the Sheppard (2006) recommendation, the classes with less than 20 students were marked as small classes 'SMALL'; classes with more than 20 and less than 30 students were noted as medium classes 'MEDIUM'; and finally the classes with more than 30 students were treated as large classes 'LARGE', and they were coded 0, 1, & 2, respectively.

Variable	Description
Level 3, Semester-Level Variables: Fall 2016, Spring 2017, Fall 2017, Spring 2018, & Fall 2018	These level 3 variables referred to the exact semester the PSTs interned. Fall 2016 was the first semester included in the study and Fall 2018 was the last. For the ease of analysis, they were coded as follow: 0 for Fall 2016, 1 for Spring 2017, 2 for Fall 2017, 3 for Spring 2018, and 4 for Fall 2018.

Outcome/Dependent Variable

Pretest Scores: Students' average pretest scores derived from students' performance on pretest/s given by the One-Plus PSTs before they taught a unit. This score was denoted by (*POST_SCR)***,** and it ranged between *0* through *100*.

Posttest Scores the outcome variable for the posttest model in this study. The PST reported posttest scores are the mean percentile scores per student. This score was denoted by (*POST_SCR)***,** and it ranged between *0* through *100*.

Analytical Strategies and Results

In hierarchical linear modeling, it is important to rescale the predictor and moderator variables to increase the interpretability of the parameters (Hahs-Vaughn, 2016; Enders & Tofighi, 2007) or to disentangle student and teacher level effects from the overall effects (Brincks et al., 2017). In general, there are three available options about centering in multilevel modeling (a) uncentered, (b) group mean-centered, and (c) grand mean-centered. The process of applying a raw metric of the variable to interpret the results is called uncentering. The effect is calculated using the average mean of all possible variables at a certain level, not considering their individual characteristics, which is known as grand mean centering. Likewise, if the means for different groups are drawn separately, especially to calculate the underlying difference among the available groups, it is called the group-mean centering (Hash-Vaughn, 2016). The decision about centering is necessitated by the research questions, basically on two grounds—with-in group and between-group variability. The within-group regression coefficient articulates the effect of the predictor variable within a given group; while the between-group regression coefficient expresses the effect of the group mean of the same variable on the group mean of the outcome variable (e.g., pre- and post-test scores). In other words, the between-group regression coefficient is just the coefficient for data that are aggregated (by averaging) to the group level (Snijders & Bosker, 2012).

To facilitate the estimation and interpretation of student- and teacher-level variables, they were group mean-centered, and the semester level variables were grand mean centered because the covariates in both levels are of substantive interest for this study. By centering the Level 1 and Level 2 variables around the group mean, this study was able to trace the difference within students, and teachers (expressed by individual demographic, and academic characteristics), and, between-group variability (expressed by group-level characteristics, e.g., EL vs. non-EL, etc.). On the other hand, the final clustering variable 'semester' does not have any covariate, thus, it is reasonable to center it around the grand mean to identify the average effectiveness of all PSTs regardless of their major, and individual characteristics. Comparative analysis of effectiveness over various semesters, in addition, allowed this study to trace the trend. These decisions about centering are consistent with the recommendations of various seminal methodological literature (Brincks et al., 2017; Hahs-Vaughn, 2016; Heck & Thomas, 2015; Snijders & Bosker, 2012; and Enders & Tofighi, 2007).

As a three-level analysis, this study used the Full Information Maximum Likelihood (FIML) estimation to be able to account for the desired number of parameters in the final model (McCoach, 2010). The statistical analyses were conducted using multiple statistical tools that include Microsoft Excel Spreadsheet, Statistical Package for the Social Sciences (SPSS) software, and Hierarchical Linear Modeling (HLM) v. 7.3. Microsoft Excel Spreadsheet was used as the platform to store the raw data, which was later cleaned and prepared to move to the first analytic platform SPSS. As the first analytical tool for this study, SPSS was used to compile, code, and prepare the data for further analyses on Hierarchical Linear Modeling (HLM) v. 7.3. The SPSS was also be used to run the data for assumption verification, comparative analysis, and MANCOVA. Furthermore, data were run in HLM version 7.3, during the model building process final pre- and posttest model, and the two-level longitudinal model.

The three-level multilevel model for this study was proposed to be estimated using a multistep manner. Starting with an unconditional model (a model without any predictors and moderators) to estimate the amount of variability within and between students, i.e., the intraclass correlation coefficient (ICC). It followed by the Level 1 model. Using an iterative process, all student-level variables were supposed to be tested one at a time with fixed intercept, fixed slope model, then random intercept, fixed slope, followed by random intercept, and random slope model consecutively. Constantly comparing Akaike Information Criteria (AIC) among subsequent models with Level 1 predictors, assumed to identify the variables that could significantly predict students' posttest scores. In addition, this study proposed a hypothesized model for the 3-level HLM. Level 1 model was executed using the following formula:

Hypothesized Model

Level 1,

$$Y_{ijk} = \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk}\, a_{pjk} + e_{ijk} \tag{5}$$

Where,

π_{pjk} (*p = 0, 1, …, p*) are Level 1 Coefficients,

a_{pjk} is a Level 1 predictor *p* for case *i* in Level 2 unit *j* and Level 3 unit *k*,

e_{ijk} is the Level 1 random effect, and

σ^2 is the variance of e_{ijk}, that is the Level 1 variance

Here, this study assumes that the random term $e_{ijk} \sim N\,(0, \sigma^2)$,

To elaborate this formula, the Level 1 design for this study will look like,

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}\,(Pre\_Scr) + \pi_{2jk}\,(EL\_status) + \pi_{3jk}\,(FRPL\_status) +$$

$$\pi_{4jk}\,(SDES\_status) + \pi_{5jk}(gender) + \pi_{7jk}\,(ethnicity) + e_{ijk} \tag{6}$$

Where,

Y_{ijk} is the posttest score for student *i* who attended the PST *j*'s classroom during the semester *k*, which is modeled as a function of intercept for each PST *j* in the semester (π_{0jk}), seven student-level predictors, and a random student-level error (e_{ijk}).

Then, Level 2 variables were introduced in the model one at a time to gauge their moderation effect in the prediction of students' learning gains. It is important to note that each of the student-level predictors becomes the outcome variable in the PST-level model. Below formulae were tested in Level 2 to identify the statistically significant predictors and moderators.

Level 2

a. Unconditional Model: Student Level Coefficient that is Fixed at the Same Value for all PST Level Units

$$\pi_{pjk} = \beta_{p0k} \tag{7}$$

b. For the Student Level Variables that do not Vary Randomly among PST Level Moderators

$$\pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} \tag{8}$$

c. For the Student Level Variables that Randomly Vary across PST Level Moderators

$$\pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} + r_{pjk} \tag{9}$$

Where,

β_{pqk} (q = 0, 1, …, Q) are PST-level coefficients,

X_{qjk} is a PST-level moderator (e.g., PST's major), and

r_{pjk} is a PST-level random effect.

Using this method, this study will be able to identify the final structure model for this study, which can be denoted by:

Level 3

a. PST Level Coefficient that is Fixed at the Same Value for all Semester Level Units

$$\beta_{pqk} = \gamma_{pq0} \tag{10}$$

b. For the PST Level Variables which do not Vary Randomly for a Semester

$$\beta_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} \tag{11}$$

c. For the PST Level Variables which Vary Randomly for the Semester

$$\beta_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} + u_{pqk} \tag{12}$$

Where,

γ_{pqs} (s = 0, 1, …, Spq) are semester-level coefficients,

W_{sk} is a semester-level variable, and

U_{pqk} is a semester-level random effect. The vector of semester-level random effects will be assumed to be distributed as multivariate normal with a mean of zero.

Finally, the classroom assessment scores are sometimes regarded as a weak measure of students' knowledge compared to standardized tests and thus are often questioned in terms of its validity and generalizability (Weber, 2009; Lavery et al., 2019; Denner et al., 2002). One of the most apparent reasons being Simpson's paradox—a statistical phenomenon of tracing false trends in multiple groups of data when analyzing groups in isolation, which either reverses or disappears once the groups are merged—that increases the chances of getting spurious regression results. This study assumed that it would be able to reduce the intercluster correlation because of the following reasons (a) a big sample size, (b) pretest scores as a covariate or predictor variable for posttest model, and (c) the result of a small-scale pilot study conducted before this project provided ample confidence to proceed with this project (Lachapelle, Oh, Shamz, Hertel, & Cunningham, 2015). On the other hand, this study acknowledged the inherent limitations of value-added measures in measuring students' achievement gain that it tends to ignore multiple critical factors such as class size, curriculum choice, home factors, socio-economic status, health, attendance, and so forth. Thus, the required cautions were practiced while analyzing the findings.

Conclusion

This chapter outlined the synopsis of the methodology of this study. Starting with the introduction, this chapter delved deep into the proposed research design and its appropriateness in answering the spelled-out research questions. Furthermore, it presented the data sources and discussed the data collection instrument in detail, including the sampling procedures and expected sample size. The data collection procedure section demarcated the variables of interest and their information, including their codes and types. This study used a non-experimental correlational research design in a three-level hierarchical nested design form, and they were executed using multiple statistical tools (e.g., Microsoft Spreadsheet, SPSS, and HLM). The analytical strategies and the results section clearly outlined the process of model building and other pertinent issues related to HLM (e.g., centering, estimation, etc.) and presented the standard models of a three-level analysis and hypothesized the generic models in different levels based on the variables of interest for this study. It further discussed the points that might threaten this study's credibility and laid out the plans to deal with them. Overall, this chapter provided complete information about the methodological aspect of this study.

CHAPTER FOUR:
RESEARCH FINDINGS

Introduction

The agenda of this correlational non-experimental quantitative study was twofold. First, assess achievement gap among student groups through the administration of pretest scores per semester for selected five semesters, and compare them with the posttest scores of the respective semester for the total change in students' learning and how it varied by students' sociodemographic characteristics. Second, calculate the rate of change in the achievement gap and teacher effectiveness by semester over the five semesters. The HLM uses the common language of regression analysis using sophisticated algorithms, structure, and lexicon, which explains its growing popularity among educational researchers. Although HLM can be estimated using maximum likelihood or Bayesian approaches, iterative computational techniques are required for either the estimation method or previously set designs (Snijders & Bosker, 2012). Starting with the unconditional (null) model, a systematic exploratory model building approach was used to identify the best fit final model with both fixed slope and random slope based on predictor variables. The significance of the predictors' effects, reliabilities, difference in the −2 Log$L$ (or chi-square) model fit statistics (i.e., $\Delta{-2}\text{Log}L = [-2\text{Log}L\text{smaller}] - [-2\text{Log}L\text{larger}]$) and the difference in the number of estimated parameters ($\Delta p = p_{\text{larger}} - p_{\text{smaller}}$) between the two models, and variance components were considered to decide the final model (Peugh & Heck, 2017; Subedi, 2005).

It is essential to note that this study tapped the standardized pretest scores as the proxy to students' prior knowledge (baseline status) of a topic in a content area course, and variability in such knowledge based on students' sociodemographic status was considered the achievement gap. The posttest scores after One-Plus PSTs' teaching intervention shed light on the rate of change in students' learning in all subjects, and grade levels, and how they vary based on PST (Level 2) and student (Level 1) characteristics. The difference in the standardized posttest scores among the students was again the achievement gap. The difference of achievement gaps between pretest and posttest models among the same student groups was considered the One-Plus PSTs effectiveness in general.

In the following section, this study will introduce two HLM models that draw inference beyond any other regression models. To firmly establish the impact of independent PST and student-level variables, this study applied maximum likelihood methods to estimate the parameters in a three-level linear model for the pretest that described the association between Level 1 variables (e.g., the students' gender, socioeconomic status, English learner status, ethnicity, and disability status), Level 2 variables (e.g., PSTs' major, the subject they taught during their internship, class size, and the grade level they taught) and Level 3 variables (the time series) or the semesters. Additional, multivariate regression (pretest and posttest scores as the dependent variables) analyses were conducted to identify the two-way interaction effects of EL variables in conjunctions with other Level 1 and Level 2 variables and their simple effects on pretest and posttest scores.

To assess the One-Plus PSTs' impact on students' test scores over time, a separate two-level HLM was modeled because this study did not find any literature on a 3-level longitudinal model with a continuous outcome. Many methodology literatures recommended the use of two-level models over three-level models when it comes to longitudinal study (e.g., Gao & Hussey, 1999). Data came from 5-different semesters starting from the Fall of 2016 until the Fall of 2018. As recommended in Hahs-Vaughn (2016), all time-invariant variables were restructured within Level 2, and time-variant variables within Level 1. As noted earlier, each PST and each student had just one wave of data, there were no time-variant variables other than the time itself. All PST and student-level data were modeled within the Level 1 and pretest and posttest scores as Level 2. The Level 1 variables included the controls for EL-status (EL), students' disability status (ESE), students' socioeconomic status (FRPL eligibility as a proxy), students' ethnicity (ETHNICITY), and gender (MALE). The PST level variables included the program in which they were admitted (MAJOR), the grade level of the students they taught during their internship, and the content area (SUBJECT) they taught.

Data Cleaning and Screening

Careful steps were taken to clean and screen the data before they were run. They were cleaned in multiple phases using the Microsoft Excel-based spreadsheet, and they were sent to Statistical Package for the Social Sciences (SPSS) for further analysis. The original data set for semesters fall 2016 through fall 2018 was composed of 809 PSTs and 21,754 students. Of all, 10 PSTs in the fall of 2016, and 8, 5, 11, and 7 during spring 2017, fall 2017, spring 2018, and fall 2018, respectively, failed to report any data on the file. It reduced the total number of PSTs down to 768 (N = 768) and students to 21,099. Again, there were a total of 290 cases of individual students who did not either have pretest or posttest statistics or the demographic information on file. Deduction of these cases brought the total number of K-12 students included in the study down to 20,809 (N = 20,809).

It is important to note that there were a total of 103 (0.49%) cases in which the PSTs did not report either the pretest or posttest scores. In such a case, the missing pre- or posttest scores were filled by the class average test score on the respective subject. Overall, there were less than 5% of missing data. The statistics of the missing data are provided in the table below. Of the total PSTs, approximately 90% provided the data for at least three pre- and posttests each, when nearly 5% of the PSTs provided the data for only one pre- and posttest. The summary statistics for sample size and missing data for both student and PST levels are presented in Table 2.

Table 2.

Sample Size and Missing Data.

SN	Semesters	PSTs	Missing at PST Level	Total PSTs (*N*)	Students	Students Missing at PST	Missing at Student Level	Total Students (*N*)
1	Fall 2016	143	10	133	4021	165	35	3821
2	Spring 2017	242	8	234	7603	120	157	7326
3	Fall 2017	100	5	95	2873	73	27	2773
4	Spring 2018	214	11	203	5054	187	60	4807
5	Fall 2017	110	7	103	2203	110	11	2083
Total		809	41	768	21, 754	655	290	20, 809

Note. PST = preservice teachers.

In addition, there was a great deal of variability on Level 2 data, especially on PSTs' major and the content area they taught. For the semester fall 2016, the PSTs came from five different majors (Elementary Education, English Language Arts Education, Math Education, Science Education, & Social Studies Education) and they taught four subject areas (Language Arts, Math, Science, and Social Studies). PSTs from all majors were present in the data of spring

2017 plus Art Education majors and the Visual Arts content area they taught. The data from the Fall of 2017 included Foreign Language Education Major PSTs as well, and all the subject areas present in the spring of 2017 plus the Foreign Language Education. The data included the TWS Graphmaker[tm] report of a total of 1 PST who majored in Foreign Language Education and taught a world language to 37 (including 4-incomplete student-level data) students during the internship. The data for the Spring of 2018 just included the PSTs from Elementary Education and Social Studies Education majors, while this study accounted for just Elementary Education major PSTs during the fall of 2018 as changes in program coordinators caused a loss of data for most secondary majors during these semesters.

Descriptive Analysis

Microsoft Excel (Spreadsheet) was used to clean the raw data, Statistical Package for the Social Sciences (SPSS) was used to calculate inferential and descriptive statistics.

The summary statistics of the student level variables are presented in Table 3. Students' English language status is the first predictor. The data collection instrument identified five subcategories of English learner variable, i.e., students who are currently enrolled in the program (LY); students who completed two years follow up (LF); students who exited the programs (LZ); Pending evaluation (LP); and not applicable or no code (ZZ or NA). After consulting the creator of the GraphMaker[TM], i.e., the data collection instrument used in this study (M. R. Lavery, personal communication, April 24, 2020) and an expert in the field (J. W. Nutta, personal communication, April 24, 2020), this study collapsed the five groups into three subgroups under *EL*. The students currently enrolled in the ESOL programs were marked as English learners.

There were a total of 1734 (8.33%) English learners (coded 1) who attended the One-Plus PSTs classes within a period of five semesters. Former ELs who completed the two years follow up and the students who exited the monitoring after the 2-year follow up program were collapsed into a single category, and it was marked *EXIT* (coded 2). A total of 704, i.e., 3.38% of students were included in this category. All remaining students were grouped under the non-English learner category (coded 0). Slightly over 88% of the total students were non-English learners. Non-English learners had the highest average pretest scores (M = 46.76, SD = 24.12) and posttest scores (M = 80.03, SD = 18.39). Students who were marked *EXIT* had lower average pretest score (M = 37.83, SD = 22.57) compared to the English learners (M = 39.23, SD = 23.04) but higher posttest scores (M = 75.44, SD = 18.85).

The predictor "FRPL" was a dichotomous variable derived from student's eligibility in the school lunch program. The students who were eligible for free or reduced-price lunch were put in the 'yes' category, which was coded 1. It is a common procedure to use his variable as a proxy for students' socioeconomic status (Furgione et al., 2018). There were a total of 9 568 students who were eligible in this program, and they represented the low-SES student population. The student population that did not qualify for the free and reduced-price lunch had higher pretest (M = 47.51, SD = 24.17) and posttest scores (M = 81.19, SD = 17.56) compared to the pretest (M = 43.86, SD = 23.90) and posttest scores (M = 76.82, SD = 20.15) of the student population that qualified for the program.

The predictor 'Ethnicity' was the reported ethnicity of the students who attended the One-Plus PSTs classes during their internship. There were a total of 8,984 White students who had an average pretest score of 47.62 and a standard deviation of 24.27, which was comparatively higher than the mean pretest scores of Black (*M* = 43.64, *SD* = 23.67) and Hispanic students (*M* = 43.16, *SD* = 23.83). Asian American (*M* = 47.95, *SD* = 24.27), American Indians (*M* = 50.47, *SD* = 23.89) and Other, or Multiracial students (*M* = 48.69, *SD* = 24.24) had higher pretest scores than White students. Posttest statistics for all ethnic groups remained in the same order except for the Other, or Multiracial groups. White students had an average posttest score of 81.47 with an s. d. of 17.35 when the average posttest scores of students in Other, or Multiracial groups was 79.60 with an s. d. of 19.31. The total number of students varied by their ethnicity. White students had the largest population (approximately 43%), while American Indians had the smallest sample size (approximately 0.4%) among all.

Students' disability status was another Level 1 predictor worthy of discussion. The GraphMaker^{TM}, the data collection instrument at the institution of study, listed a total of 18 different subcategories (e.g., speech impaired, intellectual disability, gifted, etc.) within in Students Disability Status (SWD). With the recommendation of the creator of the used version of GraphMaker^{TM}, this study reduced the subcategories into three (M. R. Lavery, personal communication, April 24, 2020). A subject area expert was consulted, and she confirmed the subcategories (J. W. Nutta, personal communication, April 24, 2020). Based on the recommendation, gifted students were kept in a separate subgroup (*GIFT*). Students without exceptionalities were grouped in 'No' subcategory, and it included students who were marked 'Not Applicable,' 'No Code,' 'Speech Impaired,' and 'Hospital Homebound' by the PSTs in the original TWS datafile. All other students were put into the student with a disability group marked 'Yes.' Of the total students, approximately 10% were students with exceptionalities, approximately 6% were gifted, and the remaining 84% were students without exceptionalities.

Gifted students were on the top of the list with the highest average pretest scores (*M* = 56.84, *SD* = 23.93) and posttest scores (*M* = 87.17, *SD* = 13.71) followed by students without exceptionalities with the average pretest scores of (*M* = 45.95, *SD* = 23.95) and posttest scores of (*M* = 79.40, *SD* = 18.66). Students with exceptionalities had the lowest average pretest (*M* = 38.17, *SD* = 22.79) and posttest scores (*M* = 72.48, *SD* = 21.45) among three groups.

Finally, the last Level 1 predictor '*MALE*' was derived from students' gender variable. Male students were put in the 'Yes' categories and females in the 'No' categories. Male students made 49.70 % of total students population; and they had slightly smaller average pretest (*M* = 45.38, *SD* = 24.48) and posttest scores (*M* = 78.67, *SD* = 19.25) compared to the average pretest (*M* = 46.18, *SD* = 23.97) and the posttest scores (*M* = 79.78, *SD* = 18.56) of the female students.

Table 3.

Descriptive Statistics for Student Level Variables.

Description	Variable	*N*	*M* (*SD*) *PRE_SCR*	*M* (*SD*) *POST_SCR*
English Language Status	*EL*			
English Learners	*Yes*	1734	39.23(23.04)	71.65(22.36)
Non-English Learners	*No*	18371	46.76(24.12)	80.03(18.39)
Students who exited the ESOL program	*EXIT*	704	37.83(22.57)	75.44(18.85)
Free or Reduced-Price Lunch (FRPL) Status	*FRPL*			
FRPL eligible students	*Yes*	9568	43.86(23.90)	76.82(20.15)
FRPL non-eligible students	*No*	11240	47.51(24.17)	81.19(17.56)
Ethnicity	*ETHNICITY*			
Non-Hispanic White	*W*	8984	47.62(24.27)	81.47(17.35)
Non-Hispanic Black	*B*	3381	43.64(23.67)	76.10(20.26)
	H	6190	43.16(23.83)	76.80(19.93)

Description	Variable	*N*	*M* (*SD*) *PRE_SCR*	*M* (*SD*) *POST_SCR*
Hispanic, Latino or Latina	*A*	890	47.95(24.27)	82.60(17.47)
Asian American or Pacific Islanders	*AI*	81	50.47(23.89)	85.36(14.13)
American Indians or Alaskan Natives	*O*	832	48.69(24.24)	79.60(19.31)
Other, Multiracial				
	ESE			
Disability Status	*Yes*	2135	38.17(22.79)	72.48(21.45)
Students with exceptionalities	*No*	17357	45.95(23.95)	79.40(18.66)
Students without exceptionalities	*GIFT*	1316	56.84(23.93)	87.17(13.71)
Gifted Students				
	MALE			
Gender	*Yes*	10341	45.38(24.48)	78.67(19.25)
Male Students	*No*	10382	46.18(23.97)	79.78(18.56)
Female Students				

Note. Level 1 consists of n = 20,809 students altogether. PRE_SCR = students' pretest scores; POST_SCR = students' posttest scores.

Further screening of the data revealed that the PSTs came from seven different teacher education programs major that included Elementary Education, English Language Arts Education, Math Education, Science Education, Social Studies Education, Art Education, and Foreign Language Education. Elementary Education was by far the most popular program based on the total number of students that PSTs who interned taught (13,413 of students) during the study period. Only a small fraction of students, i.e., 0.15% of the students were taught by the Foreign Language Education majors. The One-PSTs who were majoring English Language Arts, Math Education, Science Education, Science Education, Social Studies Education and Art Education taught total of 2150, 981,1240, 2720, & 272, students, respectively.

Table 4.

Descriptive Statistics for PST Level Variables.

Description	Variable	*N*	*M* (*SD*) *PRE-SCR*	*M* (*SD*) *POST-SCR*
Major	*MAJOR*			
Elementary Education	*ELEM*	13413		
English Language Arts Education	*ELAE*	2150		
Math Education	*MATH*	981		
Science Education	*SCIE*	1240		
Social Studies Education	*SSED*	2720		
Art Education	*ARTE*	272		
Foreign Language Education	*FLDE*	33		
Subject	*SUBJECT*			
Language Arts	*LA*	5454	50.80(23.72)	78.46(19.05)
Math	*MA*	6332	43.94(25.05)	79.24(19.78)
Science	*SC*	4640	47.70(24.08)	81.01(17.75)
Social Studies	*SS*	4057	39.17(21.04)	77.78(18.60)
Visual Arts	*VA*	272	55.56(23.78)	81.07(16.72)
World Language	*WL*	33	46.82(19.88)	90.33(14.55)
Class Size	*CL_SIZE*			
Small		7039	49.58(24.11)	81.18(19.13)
Medium		4430	47.33(24.88)	79.95(19.23)
Large		9339	42.30(23.24)	77.31(18.45)
Grade Level of Teaching	*GRADE_T*			
Elementary Grades	*ELEM*	13505	48.34(24.95)	80.40(19.18)
Middle School Grades	*MID*	1872	39.46(22.81)	78.13(17.92)
High School Grades	*HI*	5431	41.80(21.35)	76.52(18.32)

Note. Level 2 consists of n = 768 preservice teachers (who taught 20,809 students) during their internship during the study period. PRE_SCR = students' pretest scores; POST_SCR = students' posttest scores.

Furthermore, the data screening uncovered that the PSTs taught six different subject matters that included Language Arts, Math, Science, Social Studies, Visual Arts, and World Language. Of all, language arts, math, science, and social studies were regular content area courses found in all five semesters. Visual art was found to be taught only during spring 2017 and fall 2017 semesters, while world language was taught only during the fall of 2017. Among the regular content areas, students in language arts classes had the highest average pretest score (M = 50.80, SD = 23.72) followed by science (M = 47.70, SD = 24.08), math (M = 43.94, SD = 25.05), and social studies (M = 39.18, SD = 21.04). However, the in terms of the posttest scores, science had the highest scores (M = 81.01, SD = 17.75) followed by Math (M = 79.24, SD = 19.78) and language art (M = 78.46, SD = 19.05). Students had lower pretest and posttest scores in social studies.

The grade level the PSTs taught during their internship was a continuous variable that ranged from K to 12. To facilitate the estimation and to make the comparison easier, this study changed this variable into a categorical variable with three subcategories, namely elementary, middle, and high following the traditional US school categorization system. If the PSTs taught students in PK-6, they were put into the elementary group; Grades 7 & 8 in the middle group and rest in high school groups. Within the study period, the One-Plus PSTs taught 13,505 elementary students, 1,872 middle schools, and 5,431 high school students. Students in the elementary grades had slightly higher average pretest (M = 48.34, SD = 24.95), and posttest scores (M = 80.40, SD = 19.18) compared to high school students' average pretest (M = 41.80, SD = 21.35), posttest (M = 76.52, SD = 18.32) and middle school students' pretest (M = 39.46, SD = 22.81), and posttest (M = 78.13, SD = 17.92) scores.

Finally, the class size was the last Level 2 variable included in this study. It was a continuous variable derived from the total number of students per class in the raw data file. Due to the considerable variability (minimum 2 to a maximum of 167) in the number of students per class, this study categorized this variable. A meta-analysis on the Class size in K-12 schools was consulted to come up with the appropriate categories. It is important to note that most recent literature considered the student-teacher ratio more scientific than the traditional class size analysis. In addition, the class size cannot be universal among all grade levels (Sheppard, 2006). Based on the discussion of multiple studies in Sheppard (2006), this study identified three different class sizes, namely small classes (with less than 20 students), medium classes (between 20 to 30 students), and large classes (more than 30 students).

A total of 7,039 students attended the One-Plus PSTs' classes that had less than 20 students, while 4, 430 and 9,339 students attended medium and large classes. Students in the small classes had highest average pretest (*M* = 49.58, *SD* = 24.11) and posttest (*M* = 81.18, *SD* = 19.31) scores. Likewise, students in medium classes had comparatively higher pretest (*M* = 47.33, *SD* = 24.88) and posttest (*M* = 79.95, *SD* = 19.23) scores compared to the average pretest (*M* = 42.30, *SD* = 23.24) and posttest scores (*M* = 77.31, *SD* = 18.45) of the students in the large classes.

Model Building, Parameter Estimation, and Hypothesis Testing

This study had to develop multiple models (at least three), i.e., pretest model, posttest model, and the longitudinal models to answer the following research questions successfully:

1. What is the rate of change of One-Plus PSTs' effectiveness in reducing the EL achievement gap over time?

    1.1. What proportion of the variance in students' academic achievement is associated with students' EL-status, FRPL-status, ethnicity, disability status, and gender?

1.2. What proportion of variance in students' academic achievement is associated with PSTs' major, the grade of teaching, the subject they teach, and the class size?

The main research questions could be answered using a longitudinal model (Phase II), while the sub-questions could be answered using the three-level HLM models (Phase I). In its first phase, this study modeled a three-level final pretest design using a systematic exploratory approach. The final pretest model provided the basis for the final posttest model, i.e., the final posttest model included all the Level 1 and Level 2 variables included in the final pretest model including the exact same fixed or random slopes for the sake of consistency and interpretability of the research findings. The only difference in the posttest model was the outcome variable (posttest score), and it included the pretest as its predictor. Afterward, a multivariate analysis was conducted to assess the simple effect of other variables (both Level 1 and Level 2) with students' English learner status and whether they had a statistically significant interactional effect in predicting pre- and posttest scores. In Phase II, a two-level longitudinal model was designed with the posttest scores as the outcome variable, and the rate of change in teacher effectiveness was assessed.

Pretest Model Building

The Null Model

At the beginning of the study, five student-level predictors, four PST level predictors, and five-time series as the Level 3 predictors were selected. The pretest model building started with a null model, i.e., the model without any predictors (or intercept as a predictor) to assess the average pretest scores among all students regardless of their characteristics. HLM v. 7.3 was used to model, design, and run the three-level pretest models (composite of all five semesters). The linear analysis at Level 1 produced the following results of the parameter estimates. Considering the continuous pretest scores (PRE_SCR) as the outcome variable, the estimated values based on HLM output, are presented in the table below.

Table 5.

Final estimation of fixed effects for the pretest null model.

Fixed Effect	Coefficient	se	df	t Ratio	p-value
Semester model for PST level intercept, ψ_0					
PST model for student-level intercept, π_{00}					
Intercept, β_{000}	47.65	1.27	4	37.49	<.001

Note. Level 1 consists of n = 20, 809 students altogether. PST = preservice teachers.

The intercept (47.65) was found to be statistically significant (p<.001) at 5% and 1% level of significance.

The estimation of Level 2 and Level 3 variance components are presented in Tables 6 and 7, respectively.

Table 6.

Final estimation of Level 2 variance components for the pretest EL model.

Random Effect	Variance Component	df	Chi-Square	p-value
Intercept, e_0	246.66	762	17303.18	<.001

Note. Level 2 consists of n = 768 preservice teachers.

Table 7.

Final estimation of Level 3 variance components for the pretest EL model.

Random Effect	Variance Component	df	Chi-Square	p-value
Intercept, r_{00}	6.20	4	23.06	<.001

Note. Level 3 consists of n = 5 semesters.

As shown in the results, both Level 2 and Level 3 variance components were found to be statistically significant (p<.001). The reliability estimates for predicting Level 2 outcome was found to be 0.767, and the reliability estimate for predicting Level 3 outcome was found to be 0.941. These high-reliability estimates suggest that the mean pretest scores based on the given samples are the highly reliable source of the true population mean.

The estimated variance components from the unconditional model were $\sigma^2 = 246.42$, $\tau_\pi = 321.53$, and $\tau_\beta = 6.20$ at student, PST, and semester levels, respectively. The intra-class correlation (ICC) was calculated using the above variance components to assess the proportion of the variance in pretest scores between-students (PSTs), within PSTs, and across semesters, as:

PST level ICC,

$$\frac{\tau_\pi}{(\sigma^2 + \tau_\pi + \tau_\beta)} = \frac{321.53}{246.42 + 321.53 + 6.20} = \frac{321.53}{574.15} = 0.5600. \tag{13}$$

Semester Level ICC,

$$\frac{\tau_\beta}{(\sigma^2 + \tau_\pi + \tau_\beta)} = \frac{6.20}{246.42 + 321.53 + 6.20} = \frac{6.20}{574.15} = 0.0107. \tag{14}$$

These statistics indicated that 56% of the variance in the pretest score was accounted for by the teachers, while roughly 1% of the variance was accounted for by the semester. It makes sense because this study does not have any other variable in its Level 3 other than the time itself. The rest of the variance [1 - (.5600 + .0107) = 0.4293] was due to the student level characteristics.

The Conditional Model

This step marked the crucial considerations in the model building process. First, Level 1 variables were inserted into the unconditional null model one after another. The slopes were fixed for the newly entered variables, and they were randomized afterward. The models were compared systematically using variance components, deviance tests, and AIC statistics, which helped this study to decide whether to let the Level 1 variables remain in the model or to remove them. Variables with statistically non-significant intercepts were excluded from further consideration. Variables with statistically significant variance components in Level 2 & 3 could vary randomly or not. Model building in this stage was systematic, iterative, and lengthy, which resulted in numerous models. Thus, this study provides the story of including the first Level 1 variable for the demonstration purpose.

After the null model, the ELs variables (ELs and Exit) variables were included in the model with a fixed slope. The results of this model are presented in Table 8.

Table 8.

Final estimation of fixed effects for the pretest EL model.

Fixed Effect	Coefficient	se	df	t Ratio	p-value
Semester model for PST level intercept, ψ_0					
PST model for student-level intercept, π_{00}					
Intercept, β_{000}	47.65	1.27	4	37.48	<.001
Semester model for PST level intercept, ψ_1					
PST model for student-level effect of ELS, π_{10}					
Intercept, β_{100}	-9.68	0.89	20015	-18.59	<.001
Semester model for PST level intercept, ψ_2					
PST model for student-level effect of EXIT, π_{20}					
Intercept, β_{200}	-1.76	0.76	20015	-2.36	0.018

Note. Level 1 consists of n = 20,809 students altogether. PST = preservice teachers; ELS = English learners as opposed to non-English learners; EXIT = students who already exited the ESOL program.

Based on the results, the intercept (47.65) was statistically significant at 5% and 1% level of significance. Level 1 predictor ELS (English learners) were found to have statistically significantly lower pretest scores, β_{100} EL_1 = -9.68, $t(20015) = -18.59$, $P < .001$ compared to the non-EL students (coded 0). Likewise, the EL students who already exited the ESOL programs (EXIT) had statistically significantly lower pretest scores, β_{100} $EXIT_1$ = -1.76, $t(20015) = -2.36$, $p = .018$ compared to non-ELs. The 95% confidence interval of the average pretest score was 16.87 to 78.43. These results suggest that both English learners who were currently enrolled in the program and the students who exited the program or who completed the two years follow up began the instructional units with statistically significantly lower prior knowledge than their non-

EL peers. It also suggests that there was a statistically significant baseline achievement gap between ELs and non-ELs and EXIT students and non-ELs.

Multiparameter test of model fit was conducted by calculating the Akaike Information Criteria (AIC) using the deviance statistics and the total number of parameters. The general rule of thumb suggests that the model with a lower AIC value is a better fitting model (Hahs-Vaughn, 2016).

$$AIC_{null} = D + 2P = 181265.91 + 2\ (4) = 181273.91. \tag{15}$$

$$AIC_{EL} = 180922.02 + 2\ (6) = 180,\ 934.02. \tag{16}$$

The results showed that AIC_{EL} was smaller than the $AIC_{null,}$ suggesting that the more parameterized model was a better fitting model than the null model. The results of the variance-covariance component test indicated that the EL model was statistically significantly better than the non-EL model, $\chi^2 = 200.32$, $df = 1$, $p <.001$.

In the next step, both the ELS and EXIT variables could vary around their slopes randomly. The Level 2 and Level 3 variance components were found to be statistically significant for the ELS variable at 5% and 1% significance, while the non-significance results were found for the EXIT variable. Thus, the ELS variable was confirmed to have random slopes when the EXIT variable remained with the fixed slope. Tables 9 and 10 show the estimation of Level 2 and Level 3 variance components in the final EL model.

Table 9.

Final estimation of Level 2 variance components for the pretest EL model (Final).

Random Effect	Variance Component	df	Chi-Square	p-value
Intercept, e_0	246.89	431	9325.79	<.001
ELS slope, e_1	50.13	431	619.21	<.001

Note. Level 2 consists of n = 768 preservice teachers. ELS = English learners as opposed to non-English Learners.

Table 10.

Final estimation of Level 3 variance components for the pretest EL model (Final).

Random Effect	Variance Component	df	Chi-Square	p-value
Intercept, r_{00}	6.20	4	23.09	<.001
ELS/Intercept2, r_{10}	2.68	4	11.64	0.020

Note. Level 3 consists of n = 5 semesters. ELS = English learners as opposed to non-English Learners.

Both the Level 2 and Level 3 variance components were found to be statistically significant. The reliability estimates for Level 2, i.e., predicting Level 2 outcome by the final EL model, was 0.767 and by ELS was 0.553. In addition, the reliability estimates of the current model and ELS variable to accurately predict the Level 3 outcome were 0.942 and 0.275. Finally, the correlation matrix discovered that there was a negative relationship (-0.361) between the Level 1 Predictor ELS and the Intercept suggesting that the English learners who had lower pretest scores had a much steeper slope, i.e., the rate of change in students test scores were much higher among the ELs who had lower pretest scores.

The ICC of the final EL model was compared with the ICC of the null model. The results showed that the PST level ICC for this model was 0.51, which was 9% greater than in the null model. It means adding ELS as a Level 1 predictor of pretest score, randomly varying, reduced the within teacher level variability by 9%. Thus, ELs accounted for about 9% of the student-level variance in pretest scores. The semester level ICC remained somewhat similar, i.e., (.014) suggesting that 1.4% of the variance could be accounted for by the semester. Approximately 48% of the variance was due to the student level variables.

The Final Pretest Model

The process of testing all Level 1 variables one after another and the systematic insertion of Level 2 variables yielded the final pretest model for this study. Students' gender was not a statistically significant predictor of students' pretest scores, β_{100} EL_1 = -0.105, $t(18,88) = -0.423$, $p = .672$. In other words, there was not a statistically significant difference between the pretest scores between male and female students. Thus, the variable MALE was not included in the final model. The final model through following mixed model:

$$PRE\_SCR_{mti} = \beta_{000} + \beta_{010}*SUBJECT_{ti} + \beta_{020}*CL\_SIZE_{ti} + \beta 100*BLACK_{mti} + \beta_{200}*HISPANIC_{mti} + \beta_{300}*ASIAN_{mti} + \beta_{400}*FRPL_{mti} + \beta_{410}*FRPL_{mti}*GRADE_{ti} + \beta 500*DISABLE_{mti} + \beta 510*DISABLE_{mti}*SUBJECT_{ti} + \beta_{520}*DISABLE_{mti}*CL\_SIZE_{ti} + \beta 600*GIFT_{mti} + \beta 610*GIFT_{mti}*GRADE_{ti} + \beta_{700}*ELS_{mti} + \beta_{800}*EXIT_{mti} + e_{0ti} + e_{1ti} *BLACK_{mti} + e_{2ti} *HISPANIC_{mti} + e_{4ti} *FRPL_{mti} + e7_{ti} *ELS_{mti} + r_{00i} + r_{70i} *ELS_{mti} + \varepsilon_{mti} \quad (17)$$

As can be seen in the mixed model, Black, Hispanic, and Asian students from student's ethnicity were included in the final model, and their interpretation should compare with that of White students because they were the comparison groups. Other ethnic groups (Other, Multiracial, and Alaskan Native) were not in the model because they were not statistically significant predictors of students' pretest scores. Likewise, the slopes for the variables, BLACK, HISPANIC, FRPL, and ELS were randomly varying in the final models, while slopes for other variables included in the model had fixed slopes.

To assess whether the effects of Level 1 variables on students’ pretest scores vary as a function of Level 2 variables, the cross-level interactions were modeled. The final pretest model showed that the subject matter the PSTs taught during the internship, and the class size statistically significantly affected the Level 2 Intercept. There was a statistically significant moderation effect between the students’ socioeconomic status and their grade level. The content area the One-Plus PSTs taught during their internship, and the class size they taught had statistically significantly moderated the effect of Students with exceptionalities (DISABLE) on their pretest scores. Furthermore, this study found that the effect of GIFT (a Level 1 predictor) on their pretest scores were statistically significantly varied by the function of the grade level the PSTs taught during their internship (GRADE, a Level 2 variable).

Finally, the final model shows that only the intercept and the ELS (English learners vs. non-English learners) were varying randomly around their slopes in Level 3.

The AIC statistics for the null model, final EL model, and the final pretest models were conducted to identify the best fit model.

$$AIC_{Final} = 180135.24 + 2\ (34) = 180,\ 203.24. \qquad (18)$$

Based on the mentioned by the formulae 3, 4, & 6, it was evident that AIC_{Final} (180.203.34) < AIC_{EL} (180, 934.02) < AIC_{null} (181273.91) suggesting that the final pretest model was the best fit model among the EL and the Null models. Furthermore, a deviance test was conducted between the final pretest model and the model before it (in which the Level 1 slopes were allowed to vary randomly) the difference in chi-square was found to be 19.11, and there was the difference of 4 parameters between these models. The *p*-value showed that the final pretest model was statistically significantly better fitting model at the 1% of significance. In addition, ICC for the final model was calculated to gauge the variance explained at Level 1, Level 2, and Level 3. Based on the results, 47% and 1.1% of the total variance in pretest scores were accounted for by PST level differences, and by the time differences, respectively. Remaining, approximately 52% variance was due to the student level variables. Thus, adding the variables included in the final model, randomly varying and fixed slope, reduced the within PST level variability by approximately 8% in pretest scores.

The estimation of Level 2 variance components is given in Table 11, and the estimation of Level 3 variance is given in Table 12. All the Level 2 and Level 3 variance components were found to be statistically significant at .001, .003, .020, and .033 levels, respectively.

Table 11.

Final estimation of Level 2 variance components for the final pretest model.

Random Effect	Variance Component	df	Chi-Square	p-value
Intercept, e_0	240.10	209	4114.54	<.001
ELS slope, e_1	51.54	211	283.009	<.001
BLACK	21.06	215	276.42	.003
HISPANIC	5.32	215	254.66	0.033
FRPL	51.54	211	298.64	<.001
Level 1, ε	298.49			

Note. Level 2 consists of n = 768 preservice teachers. FRPL = free or reduced-price lunch status; BLACK = Black students as opposed to White students; HISPANIC = Hispanic students as opposed to White students. ELS = English learners as opposed to non-English learners.

Table 12.

Final estimation of Level 3 variance components for the final pretest model.

Random Effect	Variance Component	df	Chi-Square	p-value
Intercept, r_{00}	6.20	4	23.09	<.001
ELS/Intercept2, r_{10}	2.68	4	11.64	0.020

Note. Level 3 consists of n = 5 semesters. ELS = English learners as opposed to non-English learners.

The estimation of Level 2 coefficient reliabilities, i.e., the reliabilities of the coefficient in predicting Level 2 outcomes and the estimation of Level 3 reliabilities are presented in Table 13, and 14, respectively.

Table 13.

Estimation of Level 2 reliabilities based on the final pretest model.

Level 2 outcomes	Reliability estimate
INTRCPT1/INTRCPT2,π00	0.763
ELS/INTRCPT2,π70	0.437

Note. ELS = English learners as opposed to non-English learners.

The reliabilities coefficients for predicting Level 2 outcomes were reasonably large. However, the reliabilities coefficient for predicting Level 3 outcomes had mixed results. The intercept had relatively large reliability, i.e., 0.947, followed by the Level 1 predictor ELS, 0.231, FRPL, 0.188, and Black 0.142, respectively. The variable Hispanic had a reliability estimate of less than 10, which suggested that it was not comparatively an impressive predictor of Level 3 outcomes.

Table 14.

Estimation of Level 3 reliabilities based on the final pretest model.

Level 2 outcomes	Reliability estimate
INTRCPT1,ψ0	0.947
BLACK,ψ1	0.142
HISPANIC,ψ2	0.063
FRPL,ψ4	0.188
ELS,ψ7	0.231

Note. FRPL = free or reduced-price lunch status; BLACK = Black students as opposed to White students; HISPANIC = Hispanic students as opposed to White students. ELS = English learners as opposed to non-English learners.

Case Analysis and Assumptions

Case analyses and assumptions for each level of data that were conducted to test for the validity of the inferences made from results obtained in modeling testing depend on the reasonableness of the assumptions. Multilevel modeling should meet the assumptions of linearity, normally distributed residual, homogeneity of variance or heteroscedasticity, and independence (Hahs-Vaughn, 2016).

The assumption of linearity maintains that there is a linear relationship between variables. Linearity can be tested using the scatterplots of predicted values to the residuals within the same level. Carefully attention was paid to test the assumptions linearity in Level 1, Level 2, and Level 3 after identifying the final pretest model. Figure A5.1c through A5.1h were identified to assess the identification of the outliers. No extreme outliers were found, suggesting that the linearity assumptions were met at all levels.

In addition, residual analysis for Level 1 normality was conducted by calculating the mean, standard deviation, and skewness and Kurtosis. The results of the tests are presented in

Tables A5.1 and A5.2. Based on the results of the Level 1 residuals, this study concluded that the data met the assumption of Level 1 normalcy. The Figures A5.1a and A5.1b visualize the normality of Level 1 data. In addition, separate analyses of Level 2 and Level 3 residuals were conducted. Figures, A5.2a, and A5.2b show that the normalcy Level 2 variables; and the Figure A5.3a and A5.3b, prove that the assumptions of normality in Level 3 variables were met.

Scatter diagram for residual vs. fitted value showing the homogeneity of variance for Level 1 variables are presented in the Figures A5.1e through A5.1h. Scatter diagram for residual versus fitted values showing the homogeneity of variance for Level 1 variables are presented in Figure A5.2c. Finally, the scatter diagram for residual versus fitted values showing homogeneity of variance is presented in Figure A5.3c. These figures support the homoscedasticity, and the assumption of homogeneity of variance in Level 1, Level 2, and Level 3 models.

Results of the Final Pretest Model

The results of the final model yielded the correlation matrix between the intercept and the Level 1 variables included in the model. The following matrix shows the variance-covariance structures.

INTRCPT1	1.000				
BLACK	-0.246	1.000			
HISPANIC	-0.167	0.496	1.000		
FRPL	-0.249	-0.269	0.207	1.000	
ELS	-0.290	0.029	-0.505	0.457	1.000

Correlation between pretest scores and ELS was -0.290. It suggested that the ELs with lower pretest scores were associated with higher growth compared to the ELs who started at the higher pretest scores. Likewise, the correlation of HISPANIC, FRPL, and BLACK with pretest scores were -0.167, -0.249 & -0.246, respectively. These correlation statistics suggested that in students with lower pretest scores to begin with had much steeper slopes (i.e., higher learning gains) compared to the students who started with comparatively higher pretest scores within the same groups.

The Level 1 and Level 2 variables in the final model were groups mean-centered. The results presented in Table 15 are specific to the fall of 2016 (fall 2016 coded Zero). The intercept (the overall pretest score for the final model) was found to be statistically significant, β_{F2016} = 47.69, $t(4) = 38.67$, $P < .001$. The Black students (*M = 43.64, SD = 23.67*) had statistically significantly lower pretest scores, $\beta_{BLACK} = -3.24$, $t(751) = -7.02$, $P < .001$, $d = -0.17$ compared to their White counterparts (*M = 47.62, SD = 24.27*). It suggested that Black students had a statistically significantly lower achievement level compared to the White students. However, Cohen’s d statistics suggested a small effect size (Cohen, 1988).

Table 15.

Solution for fixed effects of pretest scores analysis.

Fixed Effect	Coefficient	se	t Ratio	df	p-value
Intercept	47.6945	1.2332	38.674	4	<.001
SUBJECT	-1.6854	0.5619	-2.999	751	.003
CL_SIZE	-2.1960	0.7232	-3.037	751	.003
BLACK	-3.2392	0.4614	-7.021	751	<.001
HISPANIC	-1.3872	0.3606	-3.847	751	<.001
ASIAN	1.4231	0.6365	2.236	16,937	.025
FRPL	-3.7032	0.4660	-7.947	751	<.001
GRADE	1.7033	0.5170	3.295	751	.001
DISABLE	-7.2280	0.4923	-14.682	16,937	<.001
SUBJECT	1.1478	0.4039	2.842	16,937	.004
CL_SIZE	2.3565	0.5267	4.474	16,937	<.001
GIFT	8.7946	0.6465	13.604	16,937	<.001
GRADE	-4.1501	0.7452	-5.569	16,937	<.001
ELS	-8.5263	0.8706	-9.793	4	<.001
EXIT	-1.2977	0.7532	-1.723	16,937	0.085

Note. Estimates shown were for n = 20,809 K-12 students taught by n = 768 One-Plus PSTs over five semesters. SUBJECT = subject taught by the One-Plus PSTs during their internship; CL_SIZE = small, medium or large class size based on the number of students, BLACK = Black students as opposed to White students; HISPANIC = Hispanic students as opposed to White students; ASIAN = Asian or Pacific Islander students as opposed to White students; FRPL = free or reduced-price lunch eligible students as opposed to non-eligible; GRADE = grade level the PSTs taught, DISABLE = students with exceptionalities as opposed to the students without one; GIFT = gifted students as opposed to the students without exceptionalities; ELS = English language learners as opposed to non-ELs; EXIT = students who exited the ESOL program as opposed to the non-EL students.

Hispanic students (*M = 43.16, SD = 23.83*) as well, had statistically significantly lower pretest scores $\beta_{HISPANIC} = -1.39$, $t(751) = -3.88$, $P < .001$, $d = -0.19$ compared to the White students. However, the Asian or Pacific Islander students (*M = 47.95, SD = 24.27*) had comparatively higher pretest scores, $\beta_{ASIAN} = -1.42$, $t(16,937) = 2.24$, $P < .001$, $d = -0.014$. In both cases, the observed effect sizes were small.

Compared to the high-SES students (*M =47.51, SD = 24.17*) the low-SES students (*M = 43.86, SD = 23.90*) had statistically significantly higher pretest scores, $\beta_{LOWSES} = -3.70$, $t(751) =$

-7.95, P < .001, *d* = -0.15. In general, a student eligible for free or reduced-price lunch program had approximately 3.70 lower pretest scores compared to her classmates who did not qualify in the program.

Students' disability statuses were the statistically significant predictors of their pretest scores. Student with disability (DISABLE) (*M =38.17, SD = 22.79*) had significantly lower pretest scores, $\beta_{DISABLE}$ = -7.23, *t*(16,937) = -14.68, P < .001, *d* = -0.33, while the gifted students (*M =56.84, SD = 23.93*) had statistically significantly higher pretest scores β_{GIFTED} = 8.79, *t*(16,937) = 13.60, P < .001, *d* = 0.45 than the students without exceptionalities (*M = 45.95, SD = 23.95*). Both observed effects (Cohen's d) were medium.

Likewise, students' EL-status was a significant predictor of their pretest scores. Both groups of English learners (ELS) (*M =39.23, SD = 23.04*) had statistically significantly lower pretest scores β_{ELS} = -8.53, *t*(4) = -9.79, P < .001, *d* = -0.32 compared to non-ELs (*M =46.76, SD = 24.12*). While the students who exited the ESOL programs (EXIT) (*M =37.83, SD = 22.57*) had comparatively lower but statistically not significant pretest scores β_{EXIT} = -1.30, *t*(16,937) = -1.72, P = .085, *d* = -0.38, compared to their non-EL counterparts. Among all the groups compared in this model, the gap between the ELs and non-ELs was the biggest in terms of difference in beta-coefficient.

As can be seen in Table 16, there were moderation (interaction) effects in the case of some Level 1 predictors, which showed the change in students' pretest scores as a function of Level 2 variables.

The mean pretest scores for a student in fall 2016 (coded 0) was 47.69, and it was statistically significantly different from a zero. There was a decrease in mean pretest scores for students in other subjects (e.g., math, science, social studies) compared to Language arts (LA coded 0), and it was statistically significant decrease relative to the language arts. In other words, students in language arts, in general, had 1.68 points higher pretest scores than in other subjects. Furthermore, controlling for SUBJECT, the effect of CL_SIZE was negative, and it was statistically significantly different from a zero. The value of 2.20 represented the decrease (because the intercept has the negative value) in students' mean pretest scores as the average class size increased by 1 unit (e.g., small to medium, or medium to large).

Furthermore, the effect of FRPL on students' pretest scores were found to change as a function of the grade level (GRADE) the One-Plus PSTs taught during the internship. It is important to note that GRADE is a categorical variable with three clear demarcations (Elementary, Middle, and High). As the result showed, there was a decrease in FRPL eligible students' mean pretest scores by approximately 1.70 units as the average grade level increased by 1 unit. In other words, low-SES students in small classes (less than 20 students) had statistically significantly higher pretest scores than the students in medium classrooms (between 20-30 students).

Likewise, the pretest scores of the students with exceptionalities were as the function of both SUBJECT and CL_SIZE, and the changes were statistically significantly different from a zero. After controlling for the class size the One-Plus PSTs taught, the subject language arts impacted these students' pretest scores negatively. In other words, students with exceptionalities had approximately 1.15 points lower pretest scores in language arts compared to other subjects. In addition, students with exceptionalities in small classes had statistically significantly higher pretest scores compared to the medium classes. A unit increase in class size would cause the disabled students' pretest scores to drop by approximately 2.36 units.

Finally, the pretest scores of the gifted students were found to be declining as the grade level increased, and such a decline was statistically significantly different from zero. Based on the results, one-unit increase in average grade level could cost the 4.15-unit decline in students' pretest scores.

To summarize, the pretest model discovered that there existed achievement gaps based on students' sociodemographic characteristics. This study did not find the difference in pretest scores as a function of students' gender; thus, the final model did not include this variable. The biggest gap in students' pretest scores was discovered between ELs and their non-EL counterparts. In addition, there were some statistically significantly higher changes in students pretest scores of certain student-level predictors caused by the PST-variables.

Final Posttest Model

Based on the above discussion. This study did not go through the model building process for the posttest model. This study aimed to measure the value-added measures of the One-Plus PSTs among their student during their internship. The model building process for the post could result in the production of a different best-fitting model than the final pretest model. If this happened, the comparison would make no sense. Thus, the final posttest model was simply the replication of the final pretest model. The only difference was the insertion of the pretest score as a predictor of students' posttest scores. The results of the final posttest model are presented in Table 16.

Table 16.

Solution for fixed effects of posttest scores analysis.

Fixed Effect	Coefficient	se	t Ratio	df	p-value
Intercept	79.1397	0.7651	103.436	4	<.001
SUBJECT	1.5015	0.3523	4.261	751	<.001
CL_SIZE	-0.9813	0.4520	-2.171	751	.030
BLACK	-1.6908	0.3648	-4.635	751	<.001
HISPANIC	-1.0477	0.2852	-3.674	751	<.001
ASIAN	0.9206	0.5226	1.762	16,935	.078
FRPL	-2.2075	0.3549	-6.221	751	<.001
GRADE	0.4402	0.3747	1.175	751	.240
DISABLE	-5.6504	0.4066	-13.897	16,935	<.001
SUBJECT	-0.1542	0.3319	-0.465	16,935	.642
CL_SIZE	1.3858	0.4327	3.202	16,935	.001
GIFT	3.2284	0.5326	6.061	16,935	<.001
GRADE	-0.5717	0.6106	-0.936	16,935	.349
ELS	-4.3968	0.6170	-7.126	4	0.002
EXIT	-0.3667	0.6151	-0.596	16935	0.551
PRE_SCR	0.3073	0.0056	54.832	16,935	<.001

Note. Estimates shown were for n = 20,809 K-12 students taught by n = 768 One-Plus PSTs over five semesters. SUBJECT = subject taught by the One-Plus PSTs during their internship; CL_SIZE = small, medium or large class size based on the number of students, BLACK = Black students as opposed to White students; HISPANIC = Hispanic students as opposed to White students; ASIAN = Asian or Pacific Islander students as opposed to White students; FRPL = free or reduced-price lunch eligible students as opposed to non-eligible; GRADE = grade level the PSTs taught, DISABLE = students with exceptionalities as opposed to the students without one; GIFT = gifted students as opposed to the students without exceptionalities; ELS = English learners as opposed to non-ELs; EXIT = students who exited the ESOL program as opposed to the non-EL students.

As can be seen in the results, the intercept was found to be statistically significant, β_{00} = 79.14, t(4) = 103.436, P < .001 for all groups included in the model for the fall of 2016. Pretest scores was found to be the statistically significant predictor, β_{PRE_SCR} = 0.3073, t(16,935) = 54.832, P < .001, of the posttest scores. The value of 0.31 shows the positive linear relationship between pretest and posttest scores. To simplify it further, every one-unit increase in students' pretest scores was associated with an approximately 0.31-unit increase in their posttest scores.

Alike in the final pretest model, the Level 1 predictors, BLACK (*M =76.10, SD = 20.26*) and HISPANIC students (*M =76.80, SD = 19.93*) had statistically significantly lower posttest scores, $\beta_{BLACK} = -1.69$, $t(751) = -4.64$, $P < .001$, $d = -0.28$ and $\beta_{HISPANIC} = -1.05$, $t(751) = -3.67$, $P < .001$, $d = -0.25$, respectively compared to their White counterparts (*M =81.47, SD = 17.35*). There was a small effect size (Cohen, 1988). The important thing to consider is the beta coefficient, the difference of posttest scores between Black students versus White students ($\beta_{BLACKpost} = -1.69$), and Hispanic students compared to White students ($\beta_{HISPANICpost} = -1.05$) were smaller than their difference in pretest scores, ($\beta_{BLACKpre} = -3.24$) and ($\beta_{HISPANICpre} = -1.39$), respectively. On the other hand, the Asian students had statistically significant higher pretest scores than the White students. However, the posttest models showed that the Asian students (*M =82.60, SD = 17.47*) no longer had statistically higher posttest scores, $\beta_{ASIAN} = 0.92$, $t(16{,}935) = 1.76$, $P = .078$, compared to their White counterparts.

Like the pretest model, student's eligibility in the free and reduced-price lunch program was a statistically significant predictor of their posttest scores. Students who were eligible in FRPL program aka low-SES students (*M =76.82, SD = 20.15*) had statistically significantly low posttest scores, $\beta_{FRPL} = -2.21$, $t(751) = --6.22$, $P < .001$, $d = -0.23$ compared to the students who were not eligible for the program (*M =81.19, SD = 17.56*). Compared to the achievement gap in the pretest score ($\beta_{FRPLpre} = -3.70$) between low-SES and high-SES groups, the gap in the posttest score is much smaller, i.e., approximately 60% reduction in the achievement gap between pretest and posttest score of low-SES and high-SES students.

Similarly, students with exceptionalities (*M =72.48, SD = 21.45*) had statistically significantly lower posttest scores, $\beta_{DISABLE} = -2.21$, $t(751) = --6.22$, $P < .001$, $d = -0.34$, as compared to the students without exceptionalities (*M =79.40, SD = 18.66*). In addition, the gifted students (*M =87.17, SD = 13.71*) had statistically significantly higher posttest scores, β_{GIFT} = 3.23, t(16, 935) = 6.06, P < .001, d = 0.47, than the students without exceptionalities. In either case, the achievement gaps between students with exceptionalities compared with students without exceptionalities and gifted versus students without exceptionalities were smaller in the posttest model compared to the pretest model.

ELs (*M =71.65, SD = 22.36*) compared with non-ELs (*M =80.03, SD = 18.39*) had statistically significantly lower posttest scores, β_{ELS} = -4.40, t(4) = -7.13, P = .002, d = 0.41, however this gap was approximately 50% narrower than in the pretest model. Like in the pretest model, students who exited the ESOL program (*M =75.44, SD = 18.85*) did not have statistically significant posttest scores compared to their non-EL counterparts. Again, the difference in mean pre- and posttest scores were approximately 4-fold lower in the posttest model.

The mean posttest score for a student in the fall of 2016 (coded 0) was 79.14, and it was statistically significantly different from a zero. After controlling for the class sizes the PST taught, there was an increase in students' mean posttest scores for other subjects of 1.50 points, and it was statistically significant increase relative to language arts (coded 0). In addition, after controlling for the subject the PSTs taught, the effect of class size on the pretest score was negative, and it was statistically significantly different from a zero. The value of -0.98 represented the decrease in posttest scores as the class size increased by 1 unit. In other words, students in small classes had higher posttest scores compared to the students in larger classes. As with the effect of L1 predictors, the moderation effect of SUBJECT and CL_SIZE were much smaller in the posttest model than in the pretest models.

The grade level the PSTs taught had a statistically significant interactional effect on students' socioeconomic status in predicting pretest scores. The grade level the PST taught positively interacted with the student's socioeconomic status in changing the posttest score; however, it was not statistically significant. Similar statistics and relationships were seen between the DISABLE and SUBJECT, and GIFT and GRADE variables. Both pairs had a statistically significant interaction effect in predicting pretest scores but not the posttest scores. After controlling for the SUBJECT, class size had a negative effect on the posttest scores of the students with exceptionalities. The value of 1.39 represented the decrease (because the intercept is negative) in mean posttest scores of disabled students as the class size increased by 1 unit. Compared to the pretest model, the difference between the group means was smaller in posttest scores.

Further Analysis of Students' EL-Status and Its Interaction with Other Level 1 and Level 2 Variables

To understand how the students' EL status interacted with other Level 1 (e.g., FRPL-status, disability status, gender, and ethnicity) and Level 2 variables (major, subject, class size, and grade levels) and how the interaction affected their posttest scores. Multivariate analysis of covariance (MANCOVA) was conducted to determine whether there were simultaneous mean differences in ELs' pretest scores and posttest scores as the function of other Level 1 and Level 2 variables. It is important to note that the effect sizes, i.e., partial η^2, are small and they are as expected because they correspond to the two-way interactional effects (e.g., ELs*Class size, etc.). The results are summarized in Table 17.

As can be seen in the table below, the omnibus Wilks' Lambda was statistically significant interaction for the interaction between student's EL status and the PSTs' major, Λ = .997, $F(20, 41576) = 2.72$, p<.001, partial $\eta^2 < .001$; PST's subject, Λ = .998, $F(16, 41544)$ = 2.72, $p = .002$, partial $\eta^2 < .001$; grade level the PSTs taught during their internship, Λ = .999, $F(8, 41596) = 3.17$, $p = .001$, partial $\eta^2 < .001$; students' eligibility in the free or reduced-price lunch program, Λ = .999, $F(4, 40698) = 2.84$, $p = .023$, partial $\eta^2 < .000$; and students' disability status, Λ = .999, $F(8, 40698) = 3.64$, $p = .000$, partial $\eta^2 < .001$. These results suggested that there could be statistically significant achievement gap on pretest and/or posttest scores between ELs and non-ELs. The results of between subject effects shed further light in this aspect.

In contrast, the omnibus Wilks' Lambda was not statistically significant for the effect of students' EL status in terms of the effect of the grade levels the PSTs taught, Λ = .999, F(8, 41596) = 1.90, $p = .056$; in terms of the effect of students' gender, Λ = 1.000, F(4, 41430) = 1.26, $p = .653$; and in terms of the effect of minority, Λ = 1.000, $F(4, 40698) = .61$, $p = .283$. Based on these results, we could conclude that the variability in ELs' test scores was not caused by the function of their grade level, gender, or minority status.

Table 17.

Multivariate Test Results Examining the Extent that Each Interaction is Statistically Significant.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
EL * MAJOR	Wilks' Lambda	.997	2.72	20	41576	.000	.001
EL * SUBJECT	Wilks' Lambda	.998	2.37	16	41544	.002	.001
EL * GRADE	Wilks' Lambda	.999	3.17	8	41596	.001	.001
EL * CL_SIZE	Wilks' Lambda	.999	1.90	8	41596	.056	000
EL * MALE	Wilks' Lambda	1.000	.61	4	41430	.653	.000
EL * MINORITY	Wilks' Lambda	1.000	1.26	4	40698	.283	.000
EL * FRPL	Wilks' Lambda	.999	2.84	4	41602	.023	.000
EL * ESE	Wilks' Lambda	.999	3.64	8	41596	.000	.001

Note. EL = English learners; MAJOR = preservice teacher's area of study; SUBJECT = subject are the preservice teachers taught during internship; GRADE = grade level the preservice teachers taught; CL-SIZE = class size the preservice teachers taught; MALE = students' gender; MINORITY = students' minority status; FRPL = students' eligibility for free or reduced-price lunch program; ESE = students' disability status.

The test of between-subject effects, especially the 'corrected model,' i.e., the omnibus *F* test for the univariate ANOVA testing the null hypothesis that there was no mean difference in each of the dependent variable (pretest score and posttest score) based on the groups in the independent variables (e.g., EL, Gender, etc.). The profile plots of the interactional effects of students' English language status and other variables are presented in Figures A6.1a. through A6.8b (Appendix F). The results of the between-subject effects are presented in Table 18.

As can be seen in Table 19, there was a statistically significantly variable change on ELs' posttest scores as a function of One-Plus PSTs' area of study, $F(10, 20789) = 3.44$, $p = .000$. In other words, there was a statistically significant posttest score difference among ELs taught by PSTs with different majors.

Likewise, ELs had a statistically significant posttest scores, $F(8, 20788) = 2.698$, $p =$ .000, and pretest scores, $F(8, 20788) = 2.577$, $p < .001$, difference as a function of the subject matter the PSTs taught, and as a function of the grade level the PSTs taught, $F(4, 20791) = 3.58$, $p = .006$, $F(4, 20791) = 2.77$, $p = .026$, as well, respectively.

Table 18.

Tests of Between-Subject Effects from Multivariate Analysis of Covariance.

Source	DV	Type III Sum of Squares	*df*	Mean Square	*F*	*P*-value	Partial Eta Squared
EL * MAJOR	Pretest	11235.7	10	1123.57	2.030	.027	.001
	Posttest	11908.8	10	1190.88	3.436	.000	.002
EL * SUBJECT	Pretest	11990.2	8	1498.78	2.698	.006	.001
	Posttest	7219.0	8	902.38	2.577	.008	.001
EL * GRADE	Pretest	8049.5	4	2012.37	3.583	.006	.001
	Posttest	3854.3	4	963.58	2.765	.026	.001
EL * CL_SIZE	Pretest	7648.2	4	1912.05	3.397	.009	.001
	Posttest	489.9	4	122.50	.352	.843	.000
EL * MALE	Pretest	630.4	2	315.22	.548	.578	.000
	Posttest	661.9	2	330.93	.941	.390	.000
EL * MINORITY	Pretest	1388.5	2	694.23	1.208	.299	.000
	Posttest	1256.4	2	628.20	1.799	.165	.000
EL * FRPL	Pretest	4778.8	2	2389.40	4.170	.015	.000
	Posttest	1489.9	2	749.479	2.146	.117	.000
EL * ESE	Pretest	15052.2	4	3763.04	6.697	.000	.001
	Posttest	942.7	4	235.67	.684	.603	.000

Note. EL = English learners; MAJOR = preservice teacher's area of study; SUBJECT = subject are the preservice teachers taught during internship; GRADE = grade level the preservice teachers taught; CL-SIZE = class size the preservice teachers taught; MALE = students' gender; MINORITY = students' minority status; FRPL = students' eligibility for free or reduced-price lunch program; ESE = students' disability status.

Similarly, there was a statistically significantly variability in ELs pretest scores based on the class size, $F(8, 20788) = 3.397$, $p = .009$, EL students' socioeconomic status (eligibility in the free or reduced-price lunch program), $F(8, 20788) = 4.170$, $p = .015$, and their disability status, $F(8, 20788) = 6.697$, $p < .001$. However, the ELs did not have statistically significant variability due to the class size, socioeconomic status, and disability status on posttest scores. The important point here is the pretest scores of the ELs' varied based on the class size, socioeconomic status, and disability status, but they did not affect the posttest scores.

However, the results did not see any variability in ELs' pretest and posttest scores based on their gender and minority status. It means there was not any statistically significant mean difference in ELs' pretest and posttest scores between male and female students.

To test whether the statistically significant interaction between EL and other variables was due to the large sample size, this study conducted the same MANCOVA procedure with the 20% of random samples. The random samples were selected by the SPSS automatically after providing the 20% on the random sample select cases. The SPSS selected a total of 4066 students of which 3585 were non-ELs, 365 ELs, and 116 the students who exited the ESOL programs. The results of the MANCOVA procedure of 20% random samples are presented in Table 19, and the between subject effects for pre- and posttest variation are provided in Table 20.

Compared to the total sample statistics provided in Table 18, the results of the 20% random sample vary. The statistically significant interaction effect between ELs and PSTs' Major, subject they taught, grade they taught, and students' socioeconomic status were no longer realized in the 20% random sample procedure. The further analysis of the between-subject effect, there were not statistically significant change in ELs' pretest and posttest scores as a function of other characteristics mentioned above. The only difference was observed in the pretest scores of ELs caused by the students with exceptionalities. In other words, there was statistically significant pretest scores among ELs, $F(4, 4161) = 3.20$, $p = .012$, based on whether they had been identified as a student with exceptionalities or not. Finally, like the total case model, the 20% random sample model provided with the similar effect sizes. Most of the effect sizes as measured by the partial eta squared (η^2) were less than 0.06 (Cohen, 1988). These findings strengthen the findings of the 3-level HLM model, because there wasn't any statistically significant moderation effect of PST level variables in ELs' test scores.

Table 19.

20% Random Sample MANCOVA Results Examining EL Interaction with other Variables.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
EL * MAJOR	Wilks' Lambda	.994	1.33	9	4152	.158	.003
EL * SUBJECT	Wilks' Lambda	.996	1.32	14	8302	.188	.002
EL * GRADE	Wilks' Lambda	.998	1.06	8	8320	.388	.001
EL * CL_SIZE	Wilks' Lambda	.997	1.80	8	8320	.073	.002
EL * MALE	Wilks' Lambda	1.000	.202	4	8292	.937	.000
EL * MINORITY	Wilks' Lambda	1.000	.501	4	8118	.735	.000
EL * FRPL	Wilks' Lambda	1.000	.189	4	8326	.944	.000
EL * ESE	Wilks' Lambda	.997	1.65	8	8320	.106	.002

Note. EL = English learners; MAJOR = preservice teacher's area of study; SUBJECT = subject are the preservice teachers taught during internship; GRADE = grade level the preservice teachers taught; CL-SIZE = class size the preservice teachers taught; MALE = students' gender; MINORITY = students' minority status; FRPL = students' eligibility for free or reduced-price lunch program; ESE = students' disability status.

Table 20.

Tests of Between-Subject Effects from MANCOVA 20% Random Sample.

Source	DV	Type III Sum of Squares	*df*	Mean Square	*F*	*P*-value	Partial Eta Squared
EL * MAJOR	Pretest	4028.84	9	447.65	0.814	.603	.002
	Posttest	4159.74	9	462.193	1.317	.222	.003
EL * SUBJECT	Pretest	7290.9	7	1041.56	1.915	.063	.003
	Posttest	2315.0	7	330.72	.935	.478	.002
EL * GRADE	Pretest	1436.93	4	359.23	.645	.631	.001
	Posttest	1624.31	4	406.08	1.154	.329	.001
EL * CL_SIZE	Pretest	4484.8	4	1121.18	2.014	.090	.002
	Posttest	1242.9	4	310.74	.885	.472	.001
EL * MALE	Pretest	62.03	2	31.02	.054	.947	.000
	Posttest	266.2	2	133.11	.375	.687	.000
EL * MINORITY	Pretest	807.80	2	403.90	.716	.489	.000
	Posttest	439.71	2	219.86	.626	.535	.000
EL * FRPL	Pretest	367.53	2	183.76	.323	.724	.000
	Posttest	52.71	2	26.35	.074	.928	.000
EL * ESE	Pretest	7147.1	4	1786.77	3.202	.012	.003
	Posttest	310.6	4	77.66	.222	.926	.000

Note. EL = English learners; MAJOR = preservice teacher's area of study; SUBJECT = subject are the preservice teachers taught during internship; GRADE = grade level the preservice teachers taught; CL-SIZE = class size the preservice teachers taught; MALE = students' gender; MINORITY = students' minority status; FRPL = students' eligibility for free or reduced-price lunch program; ESE = students' disability status.

Two-Level Long-Term Trend Model

An unstructured model was deemed appropriate for long-term trend analysis because of the unequal interval spacing, data collection schedules, and procedures that differed across the PSTs. Basically, the PSTs who collected the data used individual discretion and their own data collection methods using self-made instruments. During the data cleaning and screening procedures, this study witnessed a wide range of variability in the number of pre- and posttest they gave, the number of items, and their levels of difficulty, let alone the instrument they used. Such customized data collection instruments and schedules could be better understood using the time unstructured model (Hash-Vaughn, 2016). As the rule of thumb, in any hierarchical modeling, uncentering works fine if the intervals between values are equally spaced (Enders & Tofighi, 2007). The semesters in this data set were coded on a scale of 0-4 (zero representing fall 2016, while four represented fall 2018), and they were equally spaced.

This long-term modeling focused on assessing the rate of change on students' pretest score (achievement gap) and posttest score (the value-added effectiveness of One-Plus PSTs) over time. It is important to note that an individual PST and their students did not have more than one wave of data collection. The semesters represented the waves, and there were a total of five waves. Each semester had different populations and sample sizes. The following research question guided the modeling of this design and its analyses:

“What is the rate of change of One-Plus PST’s effectiveness in reducing the EL achievement gap over time?”

This question was then broken down into a few targeted questions:

a. How did the achievement gap (pretest scores) change over time?

This question could be answered using the statistics presented in Table 21. The overall model reported as the final pretest model represented the semester of fall 2016. The interpretation above did not include the rate of change in students’ pretest scores over time. The waves had been dummy coded into four variables, i.e., S2017 (spring 2017), F2017 (fall 2017), S2018 (spring 2018), and F2018 (fall 2018) and fall 2016 was a reference category. Thus, the S2017 compares the results between fall 2016 and spring 2017, F2017 compares the results between fall 2016 and fall 2017, and so on. The time variables were included in the three-level hierarchical model one after another.

Table 21.

Comparison between pretest scores by semester from 3-level HLM.

	coefficient	s e	*t*-ratio	*df*	*p*-value
ll 2016 vs. Spring 2017					
TRCPT3, β000	47.72	1.33	35.791	3	<0.001
017, β001	-5.0046	2.22	-2.254	3	0.110
ll 2016 vs. Fall 2017					
TRCPT3, β000	47.71	1.28	37.263	3	<0.001
017, β001	3.09	3.08	1.0005	3	0.389
ll 2016 vs. Spring 2018					
TRCPT3, β000	47.58	0.92	51.447	3	<0.001
018, β001	4.01	2.18	1.84	3	0.163
ll 2016 vs. Fall 2018					
TRCPT3, β000	47.74	1.17	40.673	3	<0.001
018, β001	2.19	3.10	0.729	3	0.519

Note. S2017 = spring 2017 as opposed to fall 2016; F2017 = fall 2017 as opposed to fall 2016; S2018 = spring 2018 as opposed to fall 2016; and F2018 = fall 2018 as opposed to fall 2016.

The results showed that there was a decline in students' pretest scores in the spring of 2017 ($\beta_{spring2017}$ = -5, *SE* = 2.22) compared to the fall of 2016; however, it was not a statistically significant drop. Conversely, as reported, fall 2017, spring 2018, and fall 2018 had slightly higher pretest scores ($\beta_{fall2017}$ = 3.09, *SE* = 3.08), ($\beta_{spring2018}$ = 4.01, *SE* = 2.18), and ($\beta_{fall2018}$ = 2.19, *SE* = 3.10), respectively, however, they were not statistically significant gain compared to the fall 2016. These results suggested that the achievement gap was always present between the student groups based on their sociodemographic status, and it was not declining.

b. How did the PSTs' effectiveness (posttest scores) change over time?

Like the pretest procedure, the posttest procedure asked for the change in students' posttest scores by the semester included in the study. The findings of the three-level hierarchical modeling that compared posttest scores of spring 2017, fall 2017, spring 2018, and fall 2018 to fall 2016 are provided in Table 22.

Table 22.

Comparison between posttest scores by semester from 3-level HLM.

	coefficient	s e	*t*-ratio	*df*	*p*-value
Fall 2016 vs. Spring 2017					
INTRCPT3, β000	79.20	0.64	123.588	3	<0.001
S2017, β001	0.79	1.44	0.551	3	0.620
Fall 2016 vs. Fall 2017					
INTRCPT3, β000	79.20	0.61	129.546	3	<0.001
F2017, β001	-1.17	1.60	-0.73	3	0.518
Fall 2016 vs. Spring 2018					
INTRCPT3, β000	79.20	0.54	146.692	3	<0.001
S2018, β001	2.37	1.11	2.132	3	0.123
Fall 2016 vs. Fall 2018					
INTRCPT3, β000	79.28	0.40	198.416	3	<0.001
F2018, β001	-3.14	1.13	-2.764	3	0.070

Note. S2017 = spring 2017 as opposed to fall 2016; F2017 = fall 2017 as opposed to fall 2016; S2018 = spring 2018 as opposed to fall 2016; and F2018 = fall 2018 as opposed to fall 2016.

The results showed that there was a comparatively higher posttest score among all students in spring 2017 ($\beta_{spring2017}$ = 0.79, *SE* = 1.44) and spring 2018 ($\beta_{spring2018}$ = 2.37, *SE* = 1.11) compared to fall 2016; however, they were not statistically significantly higher increase. Conversely, fall 2017 and fall 2018 had comparatively lower posttest scores compared to fall 2016, ($\beta_{fall2017}$ = -1.17, *SE* = 1.60) and ($\beta_{fall2017}$ = -3.14, *SE* = 1.13), and again, they were not statistically significantly lower posttest scores.

Based on these results, this study concluded that the One-Plus PSTs' value-added effectiveness on students' posttest scores was not declining over time.

c. Did the PSTs' effectiveness have linear or non-linear changes over time?

A separate two-level hierarchical modeling was run on HLM v. 7.0 to assess the longitudinal trend, and the results are presented in Table 23.

Table 23.

Solution for fixed effects of student and PST variables over time.

Fixed Effect	Coefficient	se	t Ratio	df	p-value
Intercept	79.8769	0.9296	85.928	748	<.001
PRE_SCR	0.3688	0.0101	36.464	742	<.001
For TIME_ID	-0.4026	0.3088	-1.304	18,870	.192
MAJOR					
ELAE	-2.0561	1.0145	-2.027	18,870	.043
MATH	-2.3006	1.2201	-1.886	18,870	.059
SCIE	-2.9142	1.6873	-1.727	18,870	.084
SSED	-1.0244	0.6019	-1.702	18,870	.089
ARTE	0.4863	1.2843	0.379	18,870	.705
FLDE	3.2754	3.6230	0.904	18,870	.366
SUBJECT					
MA	0.7778	0.3000	2.593	18,871	.010
SC	1.2909	0.3312	3.898	18,871	<.001
SS	0.1342	0.4142	0.324	18,871	.746
VA	0.7539	1.2242	0.616	18,871	.538
WL	2.3639	3.3104	0.715	748	.475
GRADE LEVEL					
MIDDLE	-0.4783	0.7065	-0.677	18,861	.498
HIGH	-4775	0.5527	-0.864	18,861	.388
CLASS SIZE					
MEDIUM	-0.0195	0.2889	-0.067	18,861	.946
LARGE	-0.7533	0.3834	-1.965	18,861	.049
ETHNICITY					
BLACK	-0.6795	0.3490	-1.947	18,861	.052
HISPANIC	-0.4443	0.3224	1.348	18,861	.168
ASIAN	0.8967	0.6656	1.348	18,861	.178
AI	-0.9400	2.1053	-0.447	18,861	.655

Fixed Effect	Coefficient	se	t Ratio	df	p-value
OTHER	-0.0279	0.4984	-0.056	18,861	.955
FRPL	-0.6586	0.22656	-2.480	18,861	0.013
ESE					
DISABLE	-0.3656	0.3109	-1.176	18,861	.240
GIFT	-0.4258	0.4903	-0.868	18,861	.385
EL	-0.0771	0.3022	-0.255	18,861	.799

Note. Estimates shown were for n = 20,809 K-12 students taught by n = 768 One-Plus PSTs over five semesters. SUBJECT = subject taught by the One-Plus PSTs during their internship; CL_SIZE = small, medium or large class size based on the number of students, BLACK = Black students as opposed to White students; HISPANIC = Hispanic students as opposed to White students; ASIAN = Asian or Pacific Islander students as opposed to White students; FRPL = free or reduced-price lunch eligible students as opposed to non-eligible; GRADE = grade level the PSTs taught, DISABLE = students with exceptionalities as opposed to the students without one; GIFT = gifted students as opposed to the students without exceptionalities; ELS = English learners as opposed to non-ELs; EXIT = students who exited the ESOL program as opposed to the non-EL students.

As can be seen in the table, the average predicted value of students' posttest scores at the beginning of the study period, i.e., fall 2016 was 79.88, and it was statistically significantly different from a zero. Students' pretest score was the only one Level 1 predictor in this model. The covariance matrix showed that lower pretest scores at the beginning of the study period were associated with higher gains.

Based on the results presented in Table 24, there was an average linear growth in students' pretest scores over five semesters, and the growth was statistically significantly different from a zero. The value of 0.37 suggested that the average rate of change in students' posttest scores for a 1-semester increase in time. Figure 5 below shows the rate of change in students' posttest scores over five semesters.

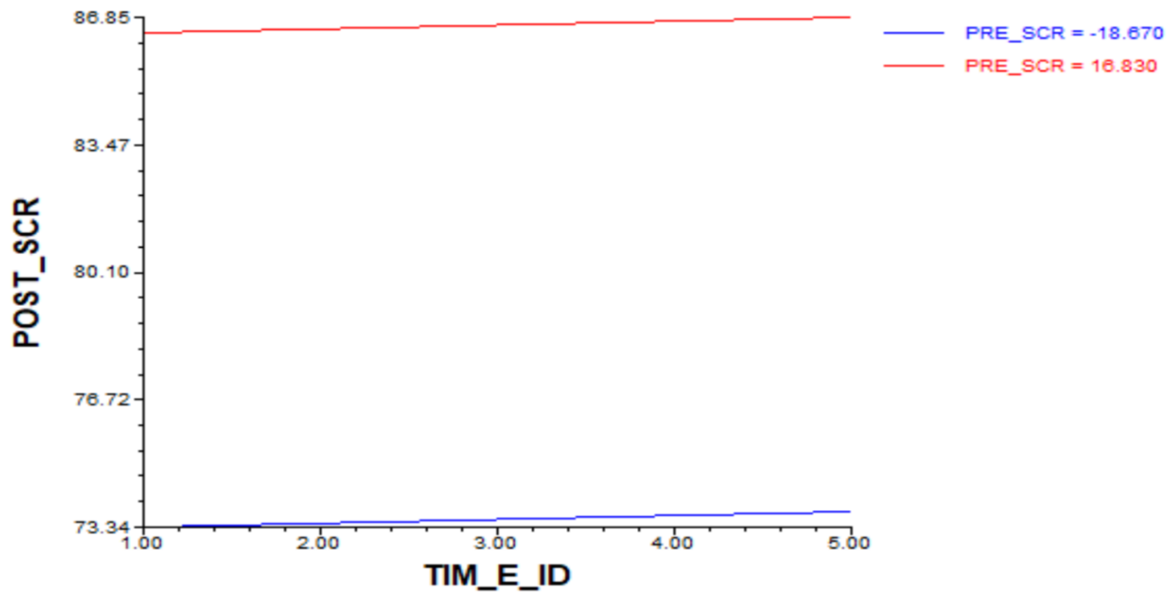


Figure 5. Time trend diagram for two pretest scores showing linear & steady but minor growth over five semesters.

Likewise, similar linear growth over time was observed among students' posttest scores in math (β_{math} = 0.78, *SE* = 0.30), and science ($\beta_{science}$ = 1.29, *SE* = 0.33). The value of 0.78 suggested that students' math posttest scores increased by 0.78 units every consecutive semester, while the value of 1.29 suggested approximately 1.29 points increase in students' posttest scores for every unit increase in time.

Conversely, some negative linear trends were assessed in students' posttest scores to certain groups of students. The average posttest scores of the students in a large classroom (more than 30 students) was statistically significantly different from a zero (β_{large} = -0.75, *SE* = 0.38). A similar trend was traced among the low-SES students. The students who were eligible for free or reduced-price lunch had approximately 0.66 unit decrease in their posttest for every unit increase in time.

The predictive capabilities of the some of the Level 1 characteristics (e.g., students EL-status, disability status, and ethnicity), including some Level 1 predictors/moderators (e.g., PST's grade level of teaching) remain unchanged for 5-semesters. Figure 6 below shows the change in students' posttest scores as a function of their minority status over time.

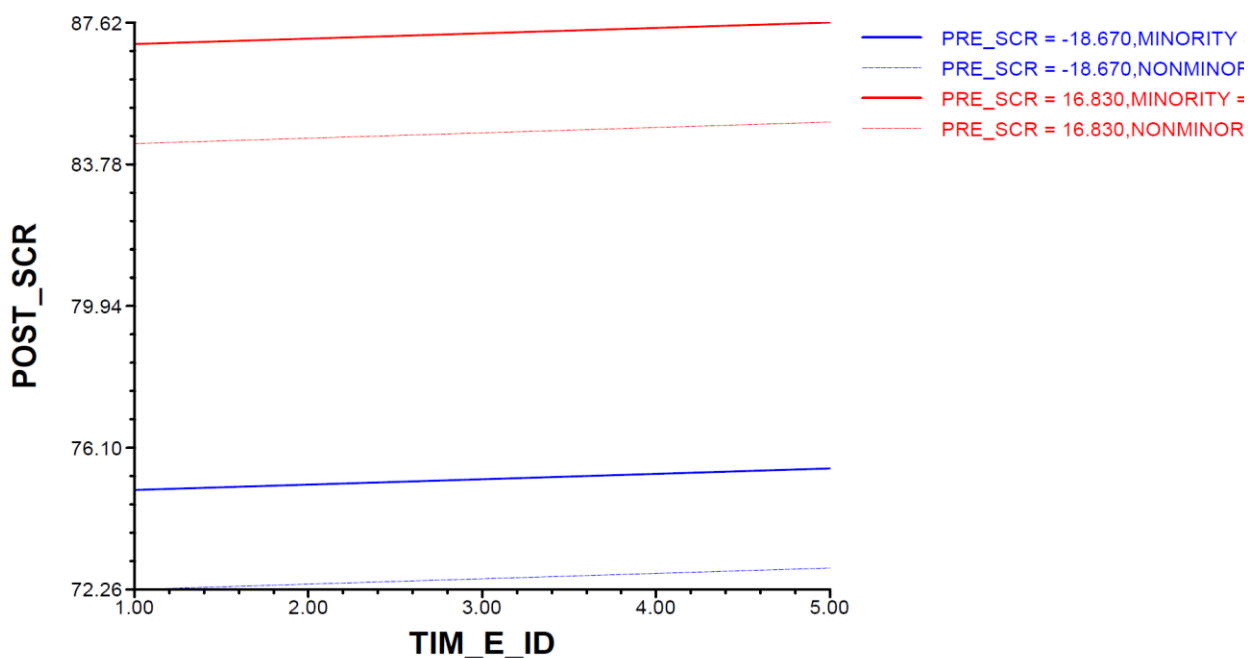


Figure 6. Time trend diagram for minority and non-minority students based on their pretest scores showing linear & steady but minor growth over five semesters.

Summary

This chapter provided a sequential demonstration of data cleaning, data screening, model building, model comparison, analysis, and evaluation of the impact of student-level predictors and One-Plus PST's characteristics on students' test scores, and how they change over time. The analyses were conducted using the PST reported data in the form of Teacher Work Sample after successful completion of their internship in any of the five semesters (e.g., fall 2016, spring 2017, fall 2017, spring 20158, or fall 2018) included in this study. The preliminary analyses of the data supported the investigations conducted to answer the research questions. Three different approaches were taken in terms of data analyses, i.e., three-level hierarchical method, MANCOVA method, and the two-level longitudinal method.

The model building process involved multiple iterative processes starting with the null, unconditional model. The Level 1 predictors were added to the model systematically one after another, and models were compared based on their ICC, AIC, and Deviance test. The variables with statistically non-significant effects were removed from the model. Upon deciding the Level 1 predictors, the Level 2 predictors were included in the model, and their moderation effect was assessed. The final model was selected as the best fitting model in terms of predictability of Level 1 predictors, an acceptable proportion of variance explained, and desirable estimated values of the coefficient (Subedi, 2005).

The results of the 3-level hierarchical model provided the answers to the two subordinate research questions.

A multivariate analysis of variance was conducted to assess all possible two-way interactional effect of students' EL status with other Level 1 and Level 2 variables, and how

ELs' test scores (both pretest and posttest) changed as the function of such variables. The results of the pretest model showed the achievement gap between student groups as the function of student-level predictors and teacher level characteristics. Based on the results, the average pretest score was 47.69 in fall 2016. Among all the Level 1 predictors, students' EL status was found to have the biggest achievement gap (-8.53) compared to their non-ELs counterparts. Students with exceptionalities had a comparatively large gap, i.e., -7.23 units compared to the students without exceptionalities. Likewise, students' achievement gap varied based on the content area and the class size. Students in larger classes (more than 30 students) had statistically significant lower pretest scores than the students in medium-sized classes (20 to 30 students), and so on.

The results of the posttest model shed light on how effective the One-Plus PSTs were in real classrooms with real students. As the result suggested, there was statically significant growth in student's posttest scores among all the students. The students with lower pretest scores had much steeper slopes compared to the student with higher scores. There were still statistically significant achievement gap among the EL and non-ELs (-4.40), low- and high-SES students (-2.21), students with exceptionalities and students without exceptionalities (-5.65), Black and White students (-1.69), and Hispanic and White students (-1.05) in the pretest model, however, the gaps were much smaller compared to the gaps in pretest model. Some gaps evident in the pretest models (e.g., Asian vs. White students) were no longer present in the posttest model. Likewise, the statistically significant moderation effect of Level 2 variable on Level 1 variable in predicting test scores (e.g., the impact of grade level on FRPL) was much narrow if not eliminated in the posttest model.

To assess the interactional effect of Level 1 and Level 2 predictors on students EL status in predicting ELs' test scores, the MANCOVA test was conducted. The multivariate test results examining the extent that each interaction based on the total student data was statistically significant which showed that students' disability status, free and reduced-price lunch status, the grade level the PSTs taught, the subject matter, and the PSTs major interacted statistically significantly with students' EL status. Further comparison of between-subject effect helped this study identify whether the significant interaction was viable in terms of predicting pre- or posttest scores or both. Comparatively, there were fewer cases of having significant interaction for posttest scores compared to the pretest scores. A separate MANCOVA was conducted using the 20% random sample on SPSS and the result showed slightly different pictures. The Statistically significant interactional effects of PSTs' major, grade, and the subject the PSTs taught, and students' socioeconomic status were no longer observed.

The comparison of average pre- and posttest scores between semester was conducted by including dummy coded time series variables as the third level predictor in the 3-level HLM model. The results showed there was not statistically significantly different in students' pretest scores over five semesters, suggesting that the achievement gaps were as big in spring 2017, fall 2017, spring 2018, and fall 2018, as they were in fall 2016. The results of the comparative posttest model provided evidence that the PST's effectiveness was consistent during these semesters. In other words, there were not any statistically significant posttest score differences between the semesters. And finally, the results of the 2-level hierarchical longitudinal model proved that there was a statistically significant linear growth in students' posttest scores over 5-semesters.

CHAPTER FIVE: CONCLUSION, DISCUSSION, AND SUGGESTIONS FOR FUTURE RESEARCH

Summary of the Study

The study was conducted using multiple analytical procedures relevant to educational and social science research as guided by the structure of the data. The 3-level hierarchical pretest model was considered the basic model which provided essential statistics of the underlying achievement gap among student groups (e.g., EL and non-ELs, Black students and White students, students with exceptionalities and students without exceptionalities, low-SES students and high-SES students) in different subjects (e.g., mathematics, science, social studies, etc.), class sizes (small, medium, and large), and grade-level (elementary, middle, or high school grades). The analysis began with the model building process geared towards identifying the best fit model for this study. Beginning with the null (unconditional) model, various conditional models were designed, starting with systematic insertion of Level 1 variables followed by Level 2 variables and constant comparison between the AIC, ICC, the deviance, and parameter test. The final pretest model was a conditional model with Level 1 and Level 2 variables with a mix of random and fixed slopes.

The final pretest model was replicated in the posttest model for the study of the rate of change in students' test scores after the treatment. The differences in average test scores between student groups from pretest and posttest provided crucial statistics that indicated the One-Plus PSTs' value-added effectiveness for this study. In addition, the results of multivariate statistics provided the answer to the question of the interactional effect of other student and PST variables with students' EL status. Finally, the results of a two-level longitudinal analysis in conjunction with the results of comparative semester wise results from the three-level pretest and posttest models enabled this study to answer the research questions.

Discussion of the Findings:

Research Question 1.1. What proportion of the variance in students' academic achievement is associated with students' EL-status, FRPL-status, ethnicity, and gender?

This study found that the average pretest scores during the fall of 2016 was 47.69; it was statistically significantly higher from a zero (p<.001). Based on the Intraclass correlation coefficient statistics for the null model, which was approximately, ICC_{Null} = .43, this study found that 43% of variance in students' pretest scores was caused for by student characteristics (EL status, ethnicity, disability status, socioeconomic status, and gender).

EL-Status

When the students' EL status was included in the model, the percentage of the variance associated with the student level characteristics was increased by approximately 12% (ICC_{EL} = .48). The change in variance was caused by the insertion of students' EL status, i.e., the total variance in students' test scores in the fall of 2016 was attributed to students' EL-status. The variance-covariance matrix for the EL model was statistically significant, suggesting that this study was missing some key student-level variables. It makes sense because this study did not use many key elements in the model, e.g., school district, school, and parent factors, including a plethora of other broader structural, systematic, and policy matters.

This study found wide achievement gaps among student groups, and the gaps were persistent throughout the pretest scores for all five semesters. Among the Level 1 predictors, the gaps between ELs and their non-EL counterparts were the biggest. Based on the findings, the ELs scored approximately 9 points lower than non-ELs, and it remained same through the study period. This finding was consistent with the prior literature on One-Plus Model. For example, Lavery et al. (2019) had found a gap of 5.32 points between EL and non-EL students, and it was second largest after the gap between students with and without exceptionalities (Lavery et al., 2019). Similarly, these results replicate the findings of the pilot study, which was conducted with smaller a sample size. Based on the results, ELs had approximately 5.05 points lower pretest scores compared to the non-ELs (Ghimire et al., 2020).

There were no differences in ELs' pretest scores based on their minority status and their gender. In other words, this study did not find any difference in the pretest scores of the male and female ELs or White and non-White ELs. However, the pretest scores of White ELs compared to non-White ELs and female ELs compared to male ELs were comparatively higher, but they were not statistically significantly higher.

*ELs*Socioeconomic status*. From further analysis of the ELs' pretest scores, this study revealed that the low-SES ELs had comparatively lower pretest scores than the high-SES ELs. As can be seen in the figure, the students who had exited the ESOL program or who completed two years follow up had lower pretest scores than ELs. The diagram below shows the students' estimated marginal mean of their pretest scores on the Y-axis and students' FRPL eligibility on they X-axis. The color-coded lines represent ELs (red), non-ELs (blue), and students who exited the ESOL program (green). We can see that there was difference in non-ELs' (blue line), ELs (red line), and exited students (green line) pretest scores based on their socioeconomic status. The student who were not eligible for FRPL program, i.e., high-SES students had comparatively higher pretest scores compared to the eligible students. The dramatic difference (steeper slope) is between the low- and high-SES students who already exited the ESOL programs. In addition, the high-SES non-ELs had the highest pretest scores, followed by the high-SES exited students, and high-SES ELs. Likewise, the low-SES non-ELS had once again the highest pretest scores, and it was followed by the low-SES ELs. The low-SES students who exited the ESOL programs had the lowest pretest scores among all the groups. The gaps between the low-SES non-ELs and the low-SES students who exited the ESOL program was the biggest. This phenomenon is interesting, and it requires further inquiry.

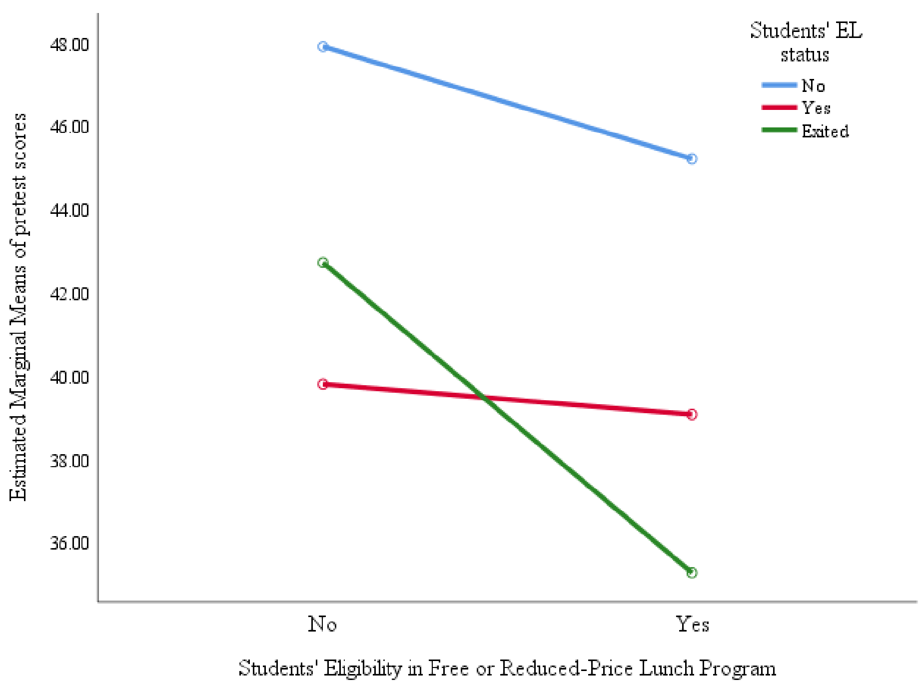


Figure 7. The profile plot of English learners' pretest scores based on their eligibility in the FRPL program.

*ELs*disability status*. In addition, students' disability status interacted statistically significantly with their EL status. ELs with exceptionalities had lower average pretest scores than the non-EL students without exceptionalities or ELs without exceptionalities. The most interesting finding in terms of students' disability status and EL status was there was an achievement gap among the EL and non-EL gifted students. As can be seen in Figure 8, gifted students who exited the ESOL program had the highest pretest scores compared to the gifted non-ELs and ELs. However, the non-ELs without any exceptionalities had comparatively higher pretest scores compared to the non-ELs with exceptionalities. Gifted students in general had higher pretest scores compared to students with and without exceptionalities. We can see the two way and three-was interaction in the figure, which was observed to be statistically significant in both MANCOVA procedures.

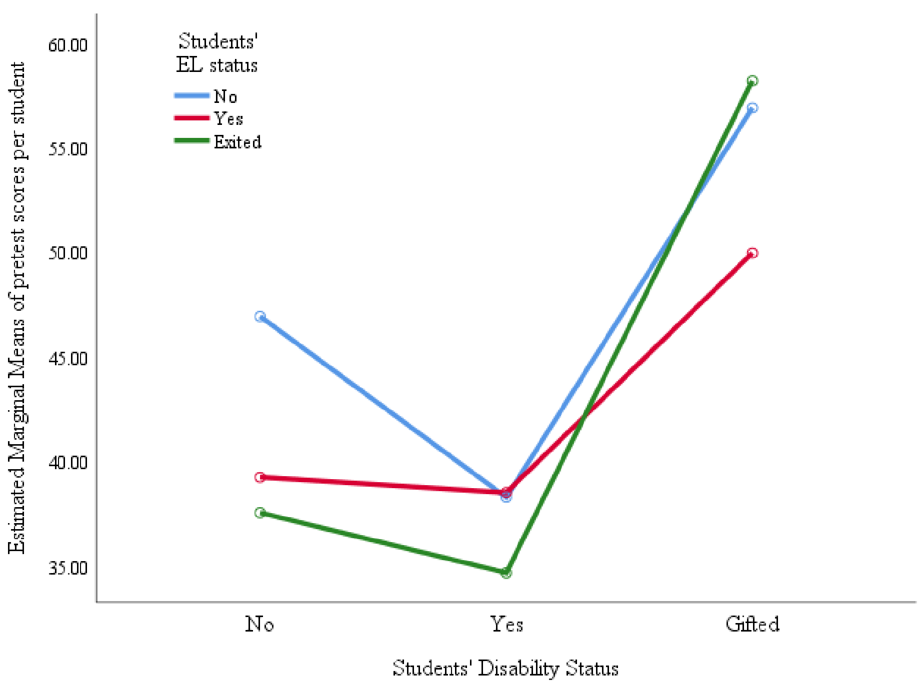


Figure 8. The profile plot of English learners' pretest scores based on their disability.

These findings were in consistent with Lohman et al. (2008) study which found a gap of approximately .5 to .6 standard deviation between EL and non-EL gifted students. A similar study by Lohman and Gambrell (2012) tested K-2 and grades 3-6 gifted students on picture verbal and picture quantitative tests. The results suggested that the gifted-ELs scored much higher in quantitative tests compared to the qualitative tests. Similarly, ELs in grades 3-6 scored much lower in English language verbal test compared to the non-verbal tests (Lohman & Gambrell, 2012). These findings hint at two critical points (a) the achievement gap between EL and non-EL gifted students is a common phenomenon, and (b) the gap seem to be a result of ELs' limited language proficiency rather than the content area knowledge (Mun, Langley, Ware, Gubbins, Siegle, Callahan, McCoach, & Hamilton, 2016). It is important to note that the data used in this study did not provide any information about the required test accommodation to students with exceptionalities. If no test accommodations were provided, then the gap should be studied further before generalizing the findings.

Likewise, the achievement gap between students with exceptionalities and students without exceptionalities students has long been studied. The findings of this study are consistent with the findings of prior studies including a longitudinal study that investigated the data for ELs with disabilities in all 50 states of the United States. The report found consistent significant gaps between exceptional students and their counterparts. Similarly, the ELs with exceptionalities had significantly lower test scores compared to ELs without exceptionalities (Thurlow, Albus, & Lazarus, 2017).

Figure 9 shows the profile plot of ELs' posttest scores based on their disability. Unlike the pretest plot, there is no interactional effect between students' EL status and disability status. This could be because of the teaching effect. As can be seen, gifted non-ELs had the highest posttest scores followed by the gifted students who exited the ESOL program. Non-ELs with exceptionalities had once again the highest possible scores followed by exited and by ELs. Among all the groups, the ELs with exceptionalities had lowest posttest scores, which were smaller than those of the ELs who had no exceptionalities or the ELs who exited the ESOL program.

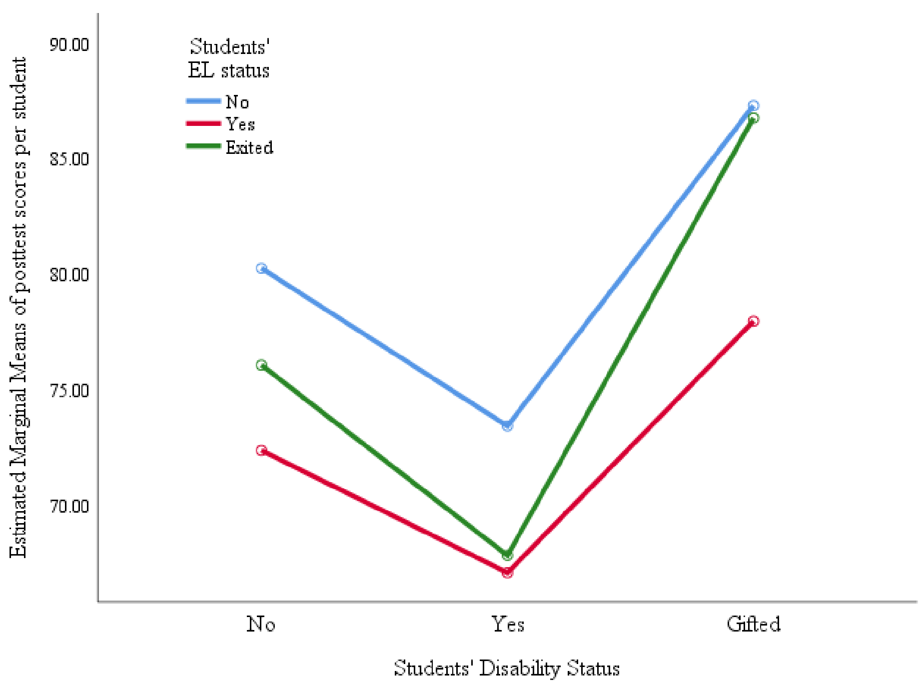


Figure 9. The profile plot of English learners' pretest scores based on their SES.

These findings are consistent with that of a report prepared by the National Center for Research on Gifted Education at the University of Connecticut that surveyed total of 45 published articles on EL and non-EL gifted students (Mun et al., 2016), which reported that ELs were not fully represented into the gifted population because of the wrong identification practices, and the linguistic barriers at the standardized testing. The Lohman et al. (2008) study concluded that there were higher chances of obtaining comparable test scores if the gifted ELs were given chances to participate in non-verbal tests compared to the verbal tests. In this context, this study noted statistically significant pretest scores between EL and non-ELs with exceptionalities, however, this gap was not statistically significant in posttest model. The One-Plus model advocates formative assessments to ELs, and it strives to provide ELs enough opportunities to be in language rich environment and participate in multiple modes of L2 communications (Nutta et al., 2012). So, the comparable results of EL and non-EL exceptional need students in posttest model, which represents the One-Plus PSTs' intervention, may be accounted for by the One-Plus model.

The primary focus of the study was the students' EL status. Thus, EL status and interactional effects were discussed separately. The impacts of all other predictors are discussed briefly.

FRPL Status, Ethnicity, Disability, and Gender

Based on the results of the pretest model, this study found that students' FRPL status, ethnicity, and disability statuses were statistically significant predictors of their pretest scores. However, there was no difference in the pretest scores of male and female students, and these findings were consistent with the prior studies (e.g., Ghimire et al., 2020; Kotok, 2017; Lavery et al., 2018; Posselt et al., 2012).

FRPL-Status. Students who were eligible for the FRPL program (low-SES students) had approximately 3.7 points lower pretest scores than their high-SES counterparts, and this difference was statistically significant. A comparable gap of 3.34 points in students pretest scores was reported by Lavery et al. (2019) study; and a gap of 3.87 points was found in Ghimire et al., (2020) study.

Students' Disability Status. Students with exceptionalities scored statistically significantly lower than the students without exceptionalities. On average, students with exceptionalities had about 7.23 points lower pretest scores compared to the students without exceptionalities, while gifted students had 8.79 points higher pretest scores compared to non-gifted students. These findings were consistent with the findings of the prior studies. Lavery et al. (2019) study found that students with exceptionalities scored 5.84 points lower than their non-exceptional counterparts, and the difference was statistically significant. Gilmour et al. (2019) conducted a metanalysis of 23 empirical studies with total of 180 effect sizes in reading between students with exceptionalities and their counterparts without exceptionalities. They reported that the students with exceptionalities read, on average, 1.17 standard deviation lower than students without exceptionalities. The study concluded that students with exceptionalities read more than 3-year below their peers without exceptinoalities.

Ethnicity. Similar to the findings of other Level 1 variables, this study found students' ethnicity to be a statistically significant predictor of their pretest scores. On average, Black students' pretest scores were 3.24 points lower, while the Hispanic students had approximately 1.39 points lower pretest scores compared to the White students. Conversely, students who were identified as Asian or Pacific Islanders had comparatively 1.42 points higher on pretest scores than White students.

Exited Students. Furthermore, the students who either already exited the ESOL program or who completed the two-year follow up had, on average, 1.30 points lower pretest scores compared to non-ELs.

Research Question 1.2. What proportion of the variance in students' academic achievement is associated with PSTs' major, the grade of teaching, the class size they teach, and the subject they teach?

PST Level Variance:

Based on the Intraclass correlation coefficient for the null model, approximately 56% (ICC_{PST} = .56) of variance in students' pretest scores is accounted for by PST characteristics. Total PST variance reduced to approximately 52% when the EL variable was included in the model. This figure is bigger than that of the students' characteristics and strengthens the assertion that teachers are instrumental in bridging the achievement gap among student groups. These findings were consistent with the prior seminal work of Darling-Hammond (2000) which reported the findings of student level variables were less influential in predicting their achievement levels compared to teacher level variables.

The findings suggested that there was a statistically significant difference in students' pretest scores in other subjects compared to language arts. On average, students in language arts had an advantage of 1.69 points over other content areas included in the study (e.g., math, science, social studies). Likewise, there was a statistically significant pretest score difference among the students who attended small (>20 students), medium (20-30 students), or large classes (more than 30 students). The value of 2.20 suggested that a unit increase in class size attributed to approximately 2.20-point decrease in students' pretest scores in that class.

The MANCOVA results of the entire sample size showed that some PST variables brought about the statistically significant differences in ELs' pretest scores. However, the results of the 20% random sample showed no statistically significant variation in ELs' pretest scores based on the subjects the PSTs taught during their internship (math, social studies, science, visual arts, and world language), the grade they taught (elementary, middle, high school grades) and the class size they taught [small (less than 20 students), middle (20-30 students), and large (more than 30 students)].

The higher average pretest scores were associated with the PSTs who taught elementary grades compared to the higher grades (e.g., elementary grades > middle grades, middle grades > high school, and elementary > high school) during their internship. Likewise, the PSTs who taught classes with fewer students saw higher pretest scores among the ELs compared to the larger class sizes. These findings suggested that ELs in higher grades and bigger classes performed much lower than the students in lower grades and smaller classes. Class sizes and grade levels are related because public schools in the states where the study took place have class size limits that increase from elementary to middle to high school. These findings were consistent with the results of a meta-analysis reported in Sheppard (2006). The profile plots for all these comparative analyses are included in Appendix F.

The pretest scores of the ELs seem to decrease by a unit increase in their grade levels. This makes sense because "the proficiency in English listening, speaking, reading, and writing is at a temporary place, a momentarily stable (yet incrementally rising) level of second language development" (Nutta et al., 2014, p. 120). As the ELs achieve the instructed target language goal, the language demands increase as the ultimate goal of an EL is to match the grade-level proficiency in reading, writing, listening, and speaking. Figure 10 below shows a visual representation of English proficiency levels and increasing language demands across grade levels.

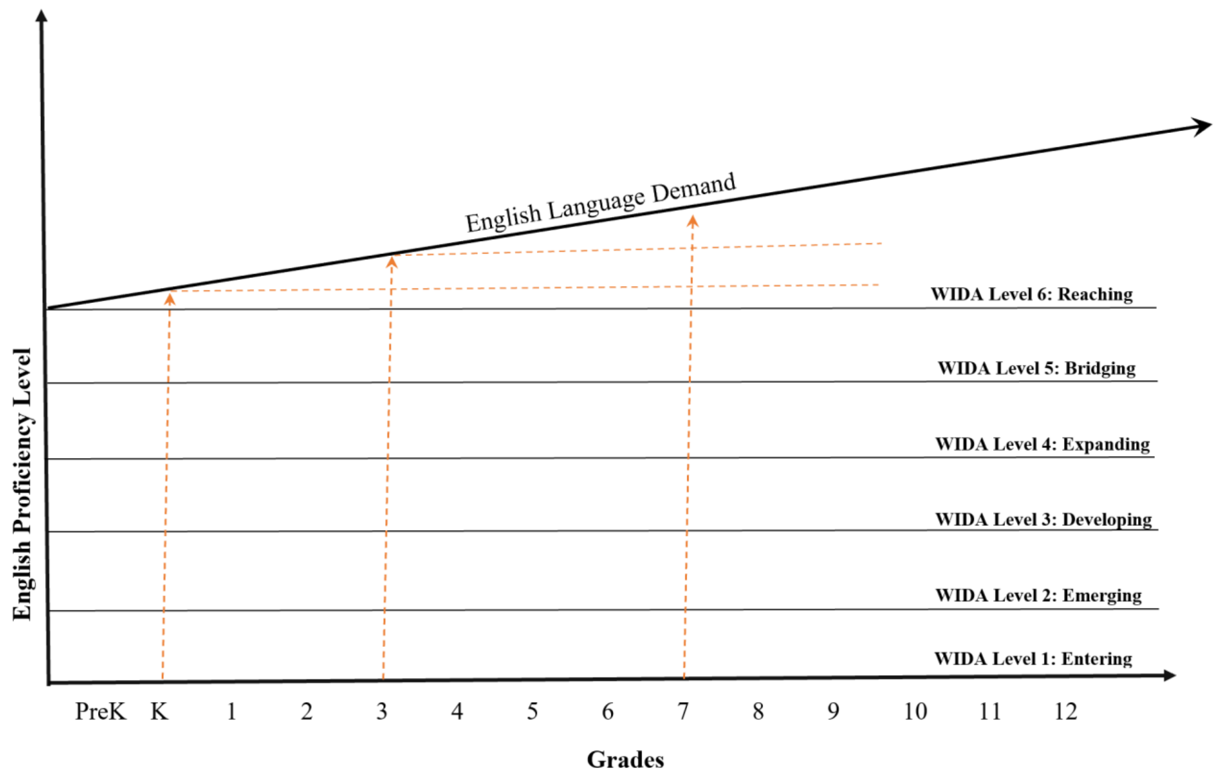


Figure 10. Diagram showing students' English proficiency level and language demand by grades. *Note*. Designed based on the discussion provided in Nutta et al., (2014)

In the figure, the X-axis shows the grade levels for ELs and the Y-axis shows the English proficiency levels as identified by the World-Class Instructional Design and Assessment (WIDA) Consortium. The vertical red arrows show the distance between the grade level language demand and WIDA Level 1 proficiency level. As the grades increase the distance is increasing, the dotted red horizontal lines show the difference in the language demand for the WIDA Level 1 students based on their grade levels. As the students' English proficiency can be the same or slightly different for the all grades, the steeper growth in language demands may help explain the bigger EL achievement gaps in higher grades than in the lower grades.

Grade level (The grade level the PSTs taught during their internship). Furthermore, the grade level the PSTs taught during their internship had a statistically significant moderation effect on students' FRPL-status in predicting their pretest scores. In other words, low-SES students and high-SES students had variation in their pretest scores based on their grade levels. In general, a unit increase in the grade level the PSTs taught (e.g., elementary to middle school grade, or middle school to high school grade), led to a disadvantage of approximately 1.70 points on students' pretest scores.

Subject (Content area the PSTs taught during their internship). Similarly, the pretest scores of the students with exceptionalities were moderated statistically significantly negatively by the subject the PSTs taught and the class size they were in. Students with exceptionalities had, on average, 1.15 points lower pretest scores on other subjects (math, science, or social studies) compared to their pretest scores in language arts. This phenomenon is similar to its effect on average pretest scores for the students in fall 2016. Likewise, a one-unit increase in the class size cost approximately 2.36 points for the students with exceptionalities in their pretest scores.

Finally, the PSTs saw differential pretest scores of gifted students, especially tied to the grades they taught. The value of -4.15 suggested that one-unit increase in the grade level (elementary to middle, or middle to high school grades) the PSTs taught, the average pretest scores of gifted students decreased by 4.15 units.

Comparison and Implications of Findings of Pretest and Posttest Models

The gap in the average pretest scores between two student groups on their pretest scores let us know the achievement gap between those student groups. This study discussed the achievement gaps between various student groups above. Now, the study of the posttest scores let us know how the students' average test scores changed after the intervention, which, in other words, is called the PSTs' value-added effectiveness (Ho & Kane, 2013; Lee, 2002). Hence, comparing the gap between student groups in pretest and posttest scores allowed this study to assess the One-Plus PSTs' effectiveness in narrowing such gaps, the fundamental motive of this study.

Based on the results, students had an average of 79.14 posttest scores in the fall of 2016. This score was statistically significantly higher than zero. Like in the pretest model, this study still found statistically significant differences in the test scores of EL and non-ELs. On average, non-ELs had a 4.40-point advantage on their posttest scores over their EL counterparts, and it was a statistically significant advantage. This difference was the second biggest only after the difference between students with exceptionalities and students without exceptionalities. However, the difference in posttest scores between EL and non-ELs was approximately 48% lower than the difference in pretest scores between these two groups.

Further analyses on ELs' posttest scores and their changes based on other characteristics had notable differences compared to the pretest model. There were statistically significant changes in ELs’ pretest scores as the functions of their FRPL-status and disability status; however, they did not bring about any statistically significant changes in students' posttest scores. These results suggested that the posttest scores of ELs were comparable between the compared student groups. Like the results in the pretest model, ELs’ posttest scores were not different based on whether they were minority or non-minority, or whether they were males or females.

Like the pretest scores, PST level variables caused a statistically significant difference in students’ posttest scores. To sum up these findings, the One-Plus PSTs had a variable impact on students’ posttest scores based on their area of study (major), the subjects they taught, the grade levels they taught, and the class size they taught during their internship.

Like in the ELs’ posttest trend, students’ eligibility in the FRPL program, disability status, and ethnicity were statistically significant predictors of posttest scores. It means there still existed statistically significant differences in posttest scores between low-SES and high-SES students, Black students and White students, Hispanic students and White students, students with exceptionalities and students without exceptionalities students, and gifted and average students. However, there were notable changes in the differences of test scores between these groups in the posttest model. The achievement gap between Black students and White students was approximately 48% lower, and between Hispanic students and White students was approximately 24% lower in posttest compared to the pretest scores. Likewise, there were 40%, 22%, and 63% reduction in test scores between low-SES students and high-SES students, students with exceptionalities and students without exceptionalities and, gifted and non-gifted students, respectively, on their posttests. There was an approximately 72% decrease in the differences of test scores between the students who exited the ESOL program or the students who completed the two-year follow up and the non-ELs from pre- to posttest scores, but this difference was not statistically significant.

In addition, there was an even greater amount of change between student groups' achievement gap in terms of the PST level variables. In the pretest model, this study identified statistically significantly higher test scores in language arts. As this study conducted the posttest model, the results showed the complete reversal of the gaps in the pretest score. Based on the findings, students' average posttest scores on other subjects (e.g., math, science, and social studies) were 1.50 units higher than in language arts, and the difference was statistically significant. We do not want to see the gaps between student groups at any level, but it is interesting to see such a drastic catapult in test scores from pre- to posttest. This study cautions that generalization of this finding may require further analysis (e.g., comparative analysis, posthoc test, etc.).

Like in the pretest model, the class size the PSTs taught during their internship had a statistically significant impact on students' average posttest scores. Based on the results, a one-unit increase in the class size (small to medium, and medium to large) would cost approximately 0.98 units decrease in their average postscores. However, the gap of 0.98 units was approximately 55% smaller than in the pretest mode.

Likewise, the pretest model found a statistically significant change in students' test scores based on their FRPL status as a function of their grade level. There was still a small negative impact of the higher grades, but it was not statistically significant. A similar trend was found on the posttest scores of students with exceptionalities and the moderation effect of the subject.

There was no difference in exceptional students' posttest scores based on the subjects the PSTs taught. Similarly, the difference in gifted students' pretest scores based on their grades no longer existed in the posttest model. The only moderation effect that remained statistically significant was disabled students' posttest scores based on the class size the PSTs taught. For every one-unit increase in their grade level, the students with exceptionalities had 1.39 points smaller posttest scores. This gap was approximately 42% lower than in the pretest model.

Research Question 1. What is the rate of change of One-Plus PST's effectiveness in reducing the EL achievement gap over time?

Using the results from the three-level pretest model, this study discovered that students' average pretest scores for spring 2017, fall 2017, spring, 2018, and fall 2018 were no different from the achievement gaps measured in the fall of 2016. They had slightly higher or lower average pretest scores, but they were not statistically significant either way. These findings are consistent with many prior long-term studies (e.g., Jeynes, 2015; Portes, 2005; Welsh, 2018) that reported stubborn achievement gaps.

As stated earlier, there was a statistically significant increase in students' posttest scores compared to the pretest scores. This study also found that higher pretest scores were associated with higher posttest scores, and ELs, on average, had lower pretest scores compared to non-ELs. There was still a statistically significant difference between EL and non-EL scores in both pretest and posttests, however, they were much smaller in their posttests. If the posttest scores are any indicators of teachers' effectiveness, there was a considerable amount of effectiveness of One-Plus PSTs' in fall 2016 as noted previously. The effectiveness remained consistent throughout the study period, i.e., there was no difference in posttest scores between the semesters fall 2016 and spring 2017; fall 2016 and fall 2017; fall 2016 and spring 2018; and fall 2016 and fall 2018. These findings were corroborated by the results of the two-level longitudinal HLM.

Based on the results of the longitudinal model, this study found that there were not any differences in students' pretest scores over time-based on their ethnicity, disability status, and their EL status compared to fall 2016. It means the results remained alike throughout the study period. There was a consistent drop in students’ posttest scores based on their socioeconomic status (FRPL), and it was statistically significant. The value of -0.66 suggested that a one-unit increase in time (semester) was associated with 0.66 points drop in students’ posttest scores.

There was a decrease of approximately 2.06 points on average posttest scores per one-unit increase in time for the PSTs who majored in English Language Arts compared to the ones who majored in Elementary Education. The PSTs who majored in Math Education, Science Education, Science Education, and Social Studies Education saw a comparatively lower posttest as the semesters progressed, but they were not statistically significant decline over time. Conversely, PSTs in Art Education and Foreign Language Development Education had comparatively higher posttest scores per unit increase in a semester, but they were not a statistically significantly increase over time.

Compared to posttest scores of language arts, PSTs who taught math and science saw significantly higher linear growth in their students’ posttest scores. Based on the findings, one-unit increase in time was associated with approximately 0.78 points increase in students’ math scores over language arts, and 1.29 points increase in science scores. The posttest scores in social studies, visual arts, and world language remained the same over time.

Compared to the average posttest scores among the students in small classrooms (with less than 20 students), students in large classes (with more than 30 students) had a statistically significant negative trend over time. Per every one-unit increase in time, the gap of average posttest scores between small class and large class increased by approximately 0.75 points. Finally, there was no change in students' average posttest scores based on the grade levels the PSTs taught and between the small and medium class sizes.

Multiple questions arise at this point. First, why is the achievement gap still present as evident in pretest model? Why was it consistent through the study period? Are we not doing anything towards eliminating it? Or as usual it is an integral aspect of the U.S. K-12 educational system? Lacking the actual data on how many of these students' classroom teachers were One-Plus graduates, we cannot know whether they are non-One-Plus teachers and may not have been prepared well; or they did not meet the professional development or certification requirements (Uro & Barrio, 2012) to work with low performing students. These huge achievement gaps exists maybe because many in-service teachers still feel that they are not adequately prepared to work with historically low performing students including ELs (Zinth, 2013); or maybe teachers are still providing the remedial curricula (Troff, 2014), or the TPP did not provide enough knowledge base and tools (Lavery et al., 2019) to work with low performing students; or maybe the schools hire alternate route teachers (Darling-Hammond, 2005); or maybe the teachers are not prepared to meet the required cultural competency (Wixon, 2015); or perhaps the mainstream content area teachers do not have adequate English language knowledge to be able to help ELs and foster equitable academic achievement. Whatever it is, based on the results, we can conclude that the great hope of narrowing let alone eliminating achievement gaps in the United States K-12 classroom (Jaynes, 2015), is not yet, realized.

Furthermore, the achievement gaps in the posttest model and its consistency over five semesters let this study know that there are still other important components that need to go into the One-Plus TPP. However, the good news is about the statistically significantly lower achievement gaps posttest models. The data came from five different semesters in which different PSTs enrolled in the semester-long internship, and they probably taught different students, however, the lower achievement gap in the posttest model remained intact, showing that it is not a fluke. As claimed, the One-Plus model seems to offer a unique opportunity for the PSTs to perceive ELs as capable of equitable learning and offer PSTs tools to appropriately assess their achievement (Nutta et al., 2012).

The lower achievement gap persisted throughout the study period. The linear upward growth (as reported in some cases) or the stable difference in achievement gaps is indeed exciting. As correlational research this study cannot confer causality, however the level of consistency might have something to do with the quality input (Hunter & Bartee, 2003) the students received from the One-Plus PSTs; or the learning opportunities the PSTs received during their teacher preparation years (Feiman-Nemsar, 2001); or the PSTs' intelligence or academic ability, subject matter knowledge, pedagogical knowledge, and teaching experience (Darling-Hammond, 2000) inbuilt in the One-Plus TPP, which were applied as effective instructional practices, fostering progress toward narrower achievement gaps (Kane & Staiger, 2002) or equity in academic achievement (Ho & Kane, 2013).

Limitations of the Study

This study is limited to the correlation of ESOL Infusion One-Plus PSTs' effectiveness in narrowing K-12 students' achievement gaps based on sociodemographic statuses (e.g., ethnicity, EL-status, disability status, FRPL-status, and gender) and PSTs' major, the subject they taught, the size of the class, and the grade-level they taught during their semester-long internship before graduating from Infused ESOL One-Plus TPP. It is important to note that this study is not an exhaustive assessment of the One-Plus model's components (e.g., service-learning, clinical experiences, micro-teaching, curriculum, etc.) and their isolated impact on establishing PSTs' knowledge base for equity education and the observation of direct implementation of such knowledge. Thus, this study is limited to the One-Plus PSTs who attended a large public university located in the Southeastern part of the United States and interned during Fall 2016, Spring 2017, Fall 2017, Spring 2018, and Fall 2018 semesters, and the pretest and posttest scores of their students.

1. Although this study found some a strong statistically significant correlation between students' posttest scores and the treatment (teaching effect), this study was not able to infer causality as it is deemed valid to all correlational research design. In addition, the big correlation statistics could be because of the large sample size [PST_N = 768, $Student_N$ = 20,809].
2. The sample size for the PST who majored the Foreign Language Development Education (FLDE) was small. Of five semesters, there was one FLDE major PST in the fall of 2017 who taught 37 students of whom 33 had enough information to be included in the study. Readers should use discretion while interpreting the findings

for this major.

3. The data collection instrument, TWS, is not free of criticism. A few studies (e.g., Darling-Hammond, 1997; Airasian, 1997; Stufflebeam 1997) have raised some concern about its generalizability and validity.
4. The pretest, posttest design further limits this study. Not having a control group to compare with the treatment group narrows its generalizability.
5. Furthermore, this study is not a conclusive document of the One-Plus PSTs' effectiveness because it looked at the impact from only one perspective, i.e., the rate of change in students' test scores after treatment and how they compared to that of prior test scores (pretest scores).
6. Furthermore, this study took the classroom assessment scores as the main component of analyses; it may not be a robust measure of students' learning and knowledge compared to standardized tests.
7. Last but not least, giving the whole credit to One-Plus PSTs for students' increased test scores may ignore many other complex and interacting factors, e.g., school district, school, and parents factors including a plethora of other larger structural, systematic, and policy matters.

Recommendation for Future Research

Despite the numerous limitations, this study provides the groundwork for several areas of importance for future research. It was a correlational study with nested data. As mentioned earlier, there have been many research aspects that have already been studied in isolation or in relation to other teacher preparation programs, which requires further inquiry from the perspective of the One-Plus model. It was a quantitative study and attempted to measure the One-Plus PSTs' effectiveness in narrowing the EL achievement gap. It did not study all possible aspects of the One-Plus model; thus, future research is required to understand the full nature of this emerging teacher preparation model.

1. This study did not have any control group. Thus, a future quantitative study measuring the One-Plus PSTs' effectiveness in narrowing EL achievement gap compared to that of non-One-Plus teachers is desirable to benchmark the results of this study.
2. As quantitative research, this study did not use PSTs' preparedness to work with ELs. A qualitative study assessing PSTs' readiness and their efficacy would allow the researchers to compare the level of readiness and or efficacy with the change in students' test scores.
3. A follow up qualitative study of the PSTs who graduated from the One-Plus program in the last five years could be conducted to assess their level of comfort in the everyday classroom and whether they think they learned something valuable as PSTs. Doing so, the researcher would be able to evaluate One-Plus graduates' knowledge and practices over time and how these changes are linked to students' academic achievement.
4. This study did not conduct the post hoc analysis of the rate of change in ELs' posttest scores as a function of other characteristics. A comprehensive post hoc analysis

would shed light on the overall aspects of EL status, which could be instrumental in designing targeted components (activities, strategies) to boost ELs' achievement.

5. A large-scale comprehensive analysis of students' sociodemographic status and how they interact with one another in predicting their academic achievement could provide insights tapping into the two-, three-, and four-way interactional effect.
6. Future research could include a delayed posttest of the students currently taught by the PSTs who graduated in the last 5-years from the One-Plus model to assess if their effectiveness has changed.
7. Various quantitative and qualitative studies on the impact of One-Plus Models' curriculum and isolated strategies embedded in the program (e.g., service-learning, simulation lab, portfolio keeping, internship, etc.) would shed light on these specifics.

Conclusion

The study provided new insights into the relationship between Infused ESOL One-Plus PSTs and their effectiveness in narrowing the EL achievement gap and the trend over five semesters. In addition, this study also assessed the relationship between students' socioeconomic status, ethnicity, gender, and disability status and how they impacted their test scores. Furthermore, the assessment of how the PST-level variables moderated the effect of student-level variables on their pretest and posttest scores let us know the fundamental underpinnings of the existing achievement gap, how they changed from semester to semester (e.g., remained the same, linear growth, linear decline). This study also measured One-Plus PSTs' effectiveness in narrowing achievement gaps (based on subject, class size, grade level, etc.) among student groups and how they changed over time.

This study identified substantial achievement gaps between students simply based on who they are (e.g., whether they come from low- and high-SES households, and if they were students with exceptionalities or not, etc.) and what happened to them from semester to semester. These findings were consistent with the prior research that studied the achievement gaps among students (e.g., Kotok, 2017; Welsh, 2018; Lavery et al., 2019; McHatton & Parker, 2013; Ghimire et al., 2020; Hunter & Bartee, 2003) strengthening that idea that the gaps persist.

In addition, this study also found that teachers have a substantial impact on students' test scores. The effectiveness, though, varies by their major, the subject they teach, the size of the class, and the grade level they teach. In general, a well-prepared teacher has higher chances of not only increasing students' test scores but eliminating the gaps among students' test scores. Those findings were similar to prior research that inquired the teachers' value-added effectiveness (e.g., Johnson & Wells, 2017; Kumpermintz, 2003; Loeb et al., 2014; Koedel & Betts, 2007; McCaffrey et al., 2009, the Met Project, 2009-2013, etc.).

Comparing these findings, this study concludes that the One-Plus PSTs were able to narrow the EL achievement gaps (persistent over time) if not eliminate them. Their effectiveness either increased over time (for some PST and student-level characteristics) or remained the same throughout the periods of five semesters. These findings strengthen the idea that the One-Plus teacher preparation model may be providing the training or elements within the program that enable PSTs to master them and to be effective in their classrooms.

APPENDIX A
INSTITUTIONAL REVIEW BOARD (IRB)
ORIGINAL APPROVAL LETTER

UCF
UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

EXEMPTION DETERMINATION

September 20, 2019

Dear Nirmal Ghimire:

On 9/20/2019, the IRB determined the following submission to be human subjects research that is exempt from regulation:

Type of Review:	Initial Study, Exempt Category
Title:	Narrowing English Learner (EL) achievement gaps: A multilevel analysis of an EL-Infused teacher preparation model
Investigator:	Nirmal Ghimire
IRB ID:	STUDY00000885
Funding:	None
Grant ID:	None

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made, and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Kamille Chap

Kamille Chaparro
Designated Reviewer

APPENDIX B
TEACHER WORK SAMPLES (TWS)

THE PROMPT SECTION (TWS) PART 1

TEACHER WORK SAMPLE 1

SEMESTER YYYY Teacher Work Sample:

A Unit on TOPIC for SETTING

FULL NAME

University of Central Florida

Figure A2.a. Teacher Work Sample (TWS), Part 1

TEACHER WORK SAMPLE 2

Contextual Factors

Begin typing your content here.

TEACHER WORK SAMPLE 3

Learning Goals/Objectives

Begin typing your expository content here.

The specific learning goals/objectives for this unit are as follows:

1. GOAL1 (State Standards: CODE).
    a. Objective(s)
2. GOAL2 (State Standards: CODE).
    a. Objective(s)
3. GOAL3 (State Standards: CODE).
    a. Objective(s)

Continue your expository content here.

Assessment Plan

Begin typing your expository content here.

Class Meetings:	Description of Assessment:	Learning Goals Assessed:
Week before unit	Pre-assessment (attached to TWS)	all
Day 2	Assessment or activity	
Day 3	Assessment or activity	
Day 4	Assessment or activity	
Day 5	Assessment or activity	
Day 6	Post-assessment (attached to TWS)	all

Continue your expository content here

Design for Instruction

Begin typing your expository content here.

Analysis of Pre-assessment Results

The unit will follow the general timeline below. For more detail, I have attached my lesson plans to this TWS.

- **Day 1:** DESCRIPTION
- **Day 2:** DESCRIPTION
- **Day 3:** DESCRIPTION
- **Day 4:** DESCRIPTION
- **Day 5:** DESCRIPTION *Add or delete days as needed.*

TEACHER WORK SAMPLE 6

Instructional Decision Making

Example One: Begin typing your content here.

Example Two: Begin typing your content here.

TEACHER WORK SAMPLE 7

Analysis of Student Learning

Begin typing your expository content here.

Paste your graphs from the TWS GraphMaker in place of this text. Use the ***default*** *paste method, either by holding the* ***Ctrl*** *key (or* ***⌘*** *key on a Mac) and hitting* ***V****, or by clicking the paste button icon on the ribbon under the* ***Home*** *tab (not the pull-down arrow).*

Continue your expository content here.

TEACHER WORK SAMPLE 8

Evaluation and Reflection

Begin typing your reflective content here.

Direction:

*Enter information **ONLY** in the cells that are shaded in blue. In order to properly sort your data to display graphs that are easier to read, perform the following steps **after** you have entered **all** student data for both the pre and post assessment: **First,** click one of the values that you entered in the **PreTotal** column and then select "Sort from Smallest to Largest" from the "Sort & Filter" menu. **Then**, click one of the values in the **PostTotal** column and select "Sort from Smallest to Largest". **Finally**, click on the tabs at the bottom of the window (i.e. **StudentGraph**, **ClassGraphs**, **SubgroupGraphs**, or **LG_Graphs**) to view the graphs created.*

Release Number:
5.1.2
Version.Update.Correction: *A change in the version number indicates a major revision, a change in the update number indicates that new features have been added, and a change in the correction number indicates that errors have been fixed.*

Candidate's Identifying Information:

Please enter information about yourself.

Candidate Information							
Four Letter TWS ID (mouse over for note)	*(B or M)*	Four Letter Program Code *(see below)*	*(yyyy)*	*(FALL or SPRING)*	Prefix (ARE, EDE, EEC, EEX, FLE, LAE, MAE, SCE, SSE)	Number (4941, 4943, 6946, 6947)	Section Number
TWSID	Degree	Program	Year	Semester	Registered Internship		Section
(ARTE, ELAE, ELEM, ERLY, EXED, FLED, MATH, SCIE, SSED) - Roll mouse over this cell to see comment.							

School Demographic Information:

Please enter information about the school at which you completed your TWS.

School Information							
Enter the percent of the students at the school that fit in each ethnic group: White, Non-Hispanic (W); Black, Non-Hispanic (B); Hispanic or Latino (H); Asian or Pacific Islander (A); or American Indian or Alaskan Native (I).					*Enter the percent of students at the school who are economically disadvantaged, English Learners, and have disabilities.*		
%White	%Black	%Hispanic	%Asian	%Am.Ind.	%EconDis	%ESOL	%SWD
Total number of students in school:							

Pre/Post Assessment Information:

Please enter the following information about your pre/post assessment measures. Your pre-assessment and post-assessment must use the same scoring information for proper analysis of student results.

Total Points Available / Mastery Level per Learning Goal					
Enter the total points available on the pre/post assessment measures for each learning goal, and the number of points that a student must earn to demonstrate matery of that learning goal. If you did not use a learning goal, leave the cells for that learning goal blank.					
	LG1	**LG2**	**LG3**	**LG4**	**LG5**
Total:					
Mastery:					

Learning Goal Content Domain				
Enter the content domain of each learning goal (e.g., LA for Reading/Language Arts, MA for Mathematics, SC for Science, SS for Social Studies - Roll mouse over this cell for complete list). If you did not use a learning goal, leave the cell for that learning goal blank.				
LG1	**LG2**	**LG3**	**LG4**	**LG5**

Figure A2.b. Teacher Work Sample (TWS), Part 2, GraphMaker™

Student Demographic and Assessment Data:

Enter information for individual students in the table below. Only students to whom you have given the same pre/post assessment measures and with whom you have used the same design for instruction should be included in your analysis.

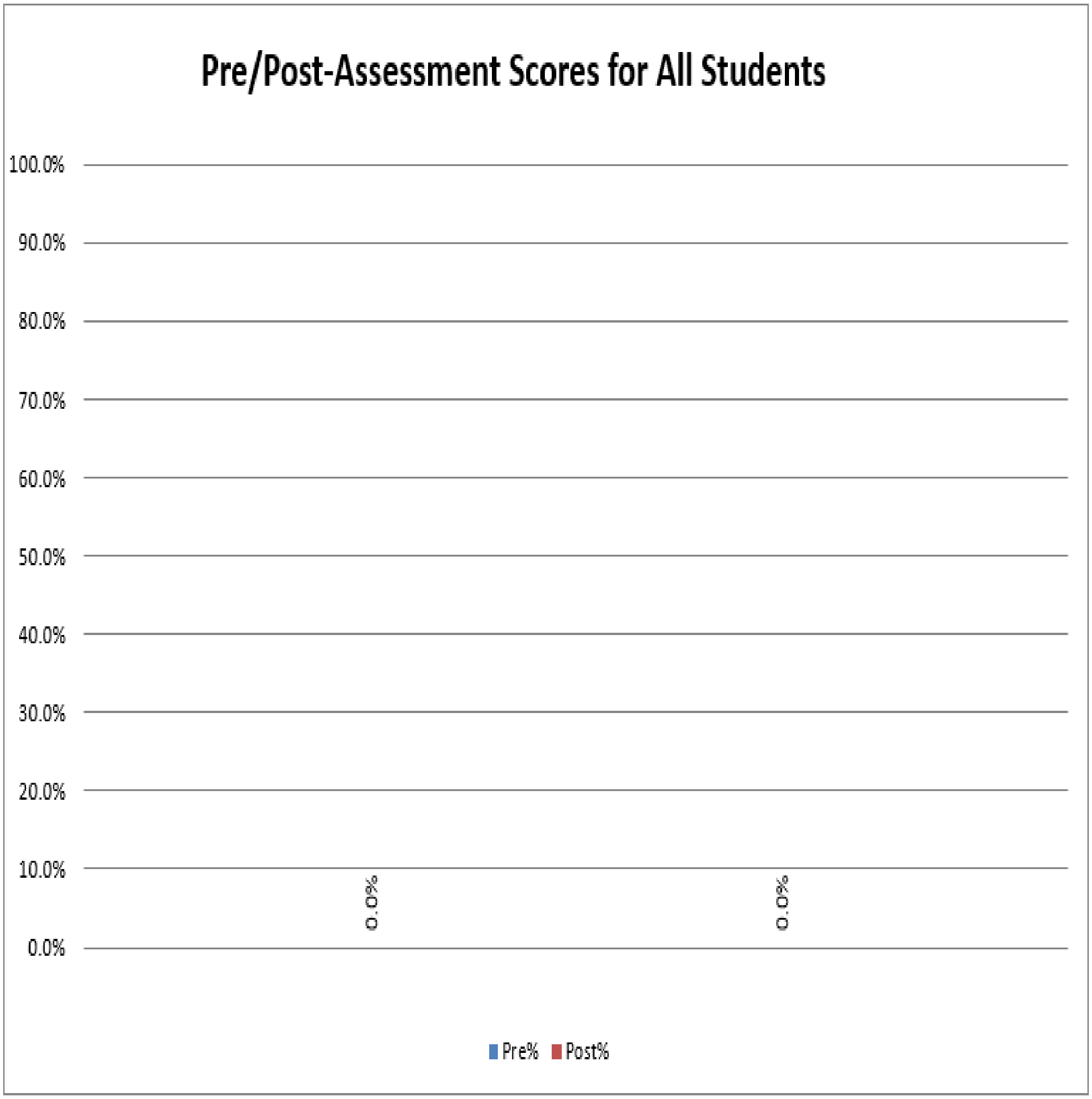
IMPORTANT: Space has been provided for up to 200 students and no one class or group may contain more than 35 students. If you have more than 200 students or classes that have more than 35 students, speak with your University Coordinator.

Demographic Information								
Anonymous Student Identifier *(must start with a letter)*	Grade Level *(PK, K, or 1 through 12)*	Class Period or Group *(single digit number 1 through 8)*	*(F or M)*	*(A, B, H, I, W, or O)*	Qualifies for free or reduced price lunch? *(Y or N)*	Primary Exceptionality *(single letter code or NA)*	ESOL Classification *(LF, LP, LY, LZ, ZZ, or NA)*	*(1*
Student	**Grade**	**Class/Grp**	**Gender**	**Ethnicity**	**EconDis**	**ESE**	**ESOL**	**O**

Pre-Assessment Data						Post-Assessment Da			
*Enter the **number of points** (not the percent correct) that students earned on the pre-assessment for each learning goal. If you did not use a learning goal, leave the cells for that learning goal blank. If a student did not take the pre-assessment, leave all scores blank. Scores that indicate mastery of the Learning Goal will be formatted in bold green*					DO NOT enter or modify data in this column.	*Enter the **number of points** (not the percent correct) t earned on the post-assessment for each learning goa use a learning goal, leave the cells for that learning g student did not take the post-assessment, leave all sco that indicate that the Learning Goal has **not** been mast*			
PreLG1	**PreLG2**	**PreLG3**	**PreLG4**	**PreLG5**	**PreTotal**	**PostLG1**	**PostLG2**	**PostLG3**	**PostLG**

Directions:

A chart showing pre/post assessment scores for individual students will be automatically generated here. ***DO NOT*** *make changes on this worksheet. Instead, copy and paste the chart into Word, then make desired changes to colors, labels, or titles. If you have more than 35 students, assign your students to a class or group on the* ***DataEntry*** *tab and use the graphs on the* ***ClassGraphs*** *tab in your TWS, instead.*

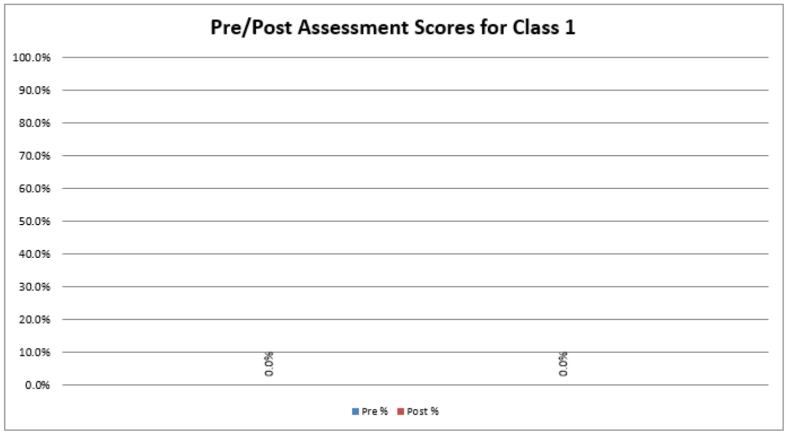


NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Directions:

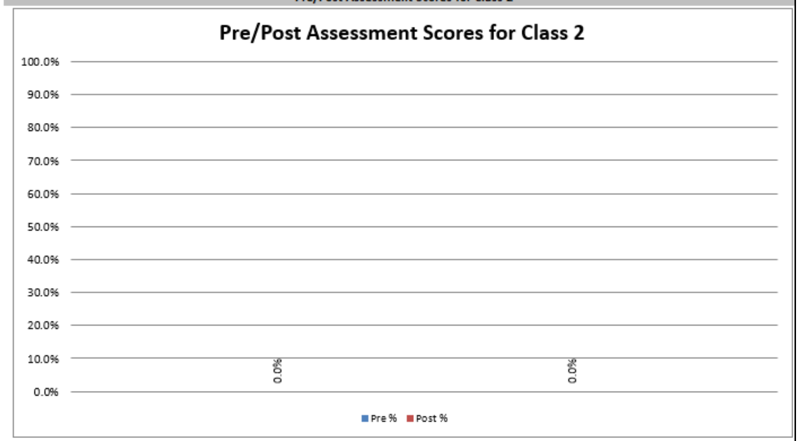
Charts Showing pre/post assessment scores for individual students in each class or group will be automatically generated here. DO NOT make changes on this worksheet. Instead, copy and paste the chart into Word, then make desired changes to colors, labels, or titles. You may need to scroll down to find all charts. Student data will be automatically sorted into the appropriate graph based on the class number (1 through 8) that you entered for them on the ***DataEntry*** *tab. No more than 35 student may be assigned to any one class or group. If you have fewer than 35 students and your students are not grouped, ignore these charts and use the graph on the* ***StudentGraph*** *tab in your TWS.*

Pre/Post Assessment Scores for Class 1



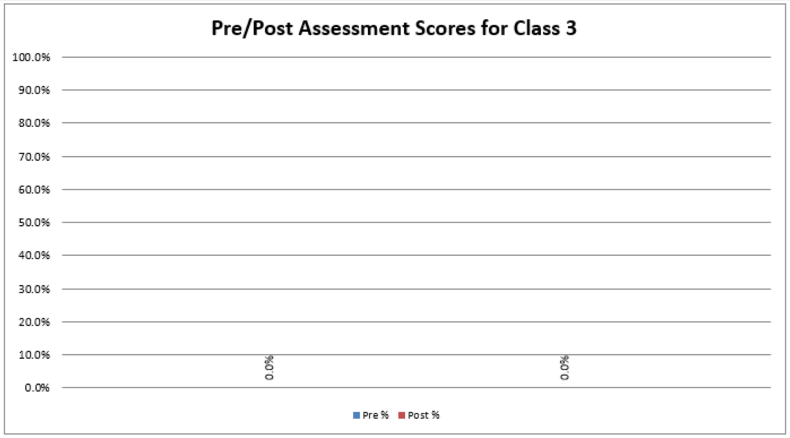
NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 2



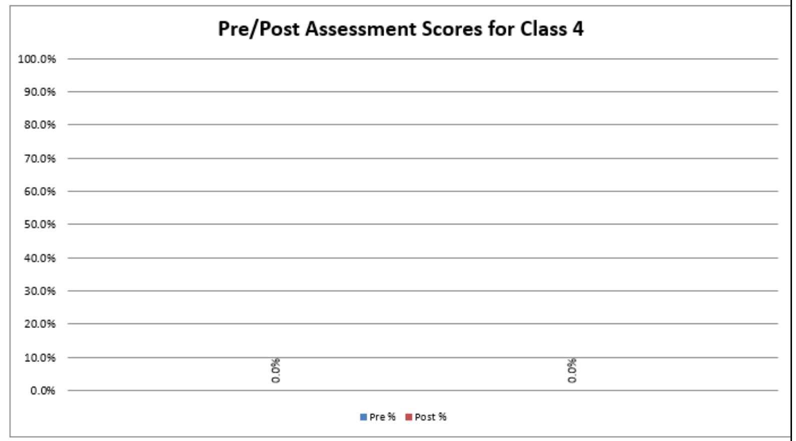
NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 3



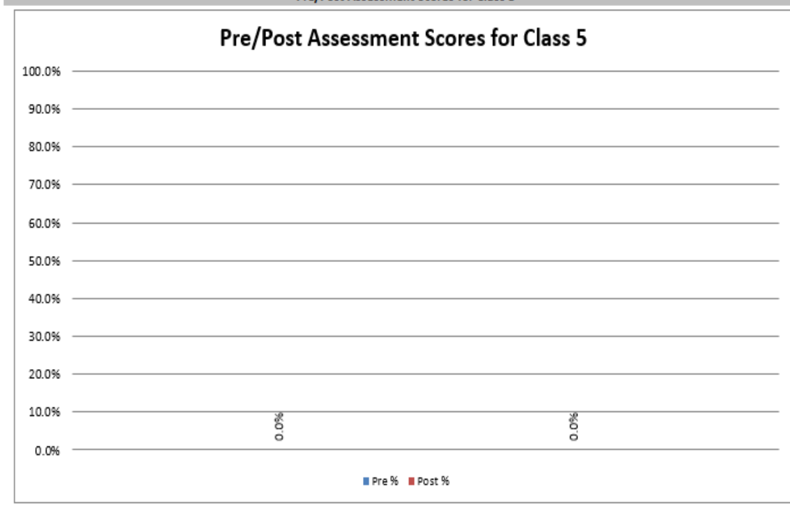
NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 4



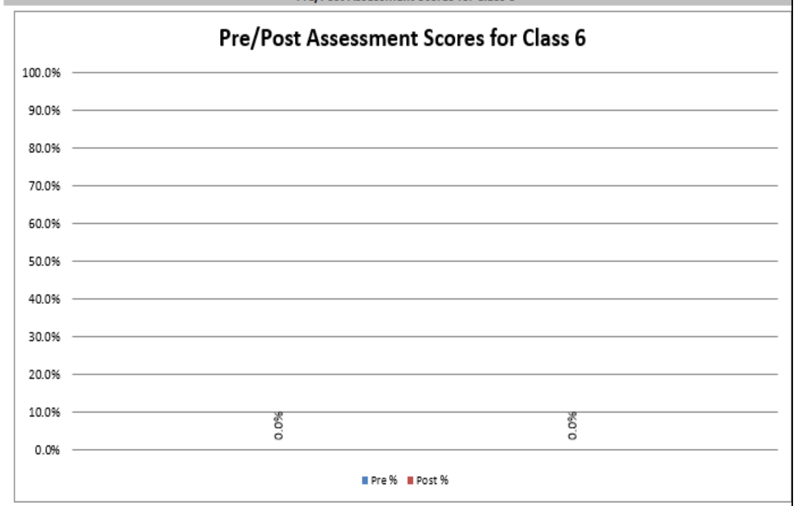
NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 5



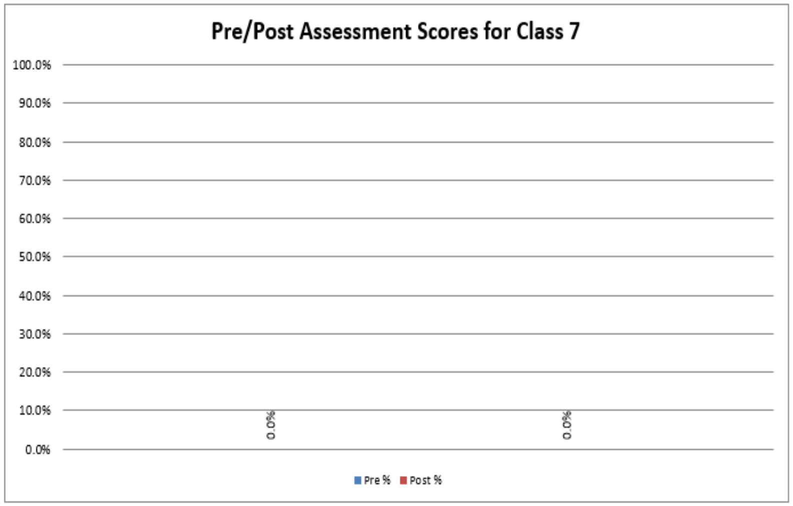
NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 6



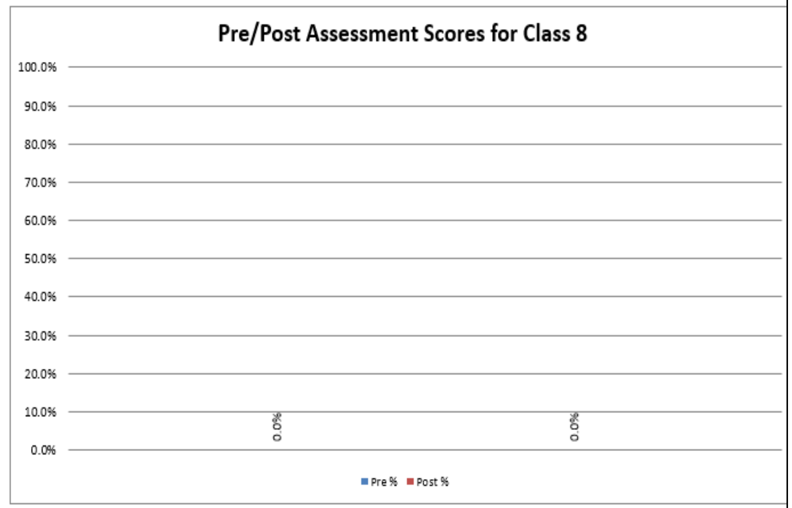
NOTE: *You will want to sort your data before using this graph in your TWS. Do this by following these two steps:* ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 7



NOTE: You will want to sort your data before using this graph in your TWS. Do this by following these two steps: ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Pre/Post Assessment Scores for Class 8

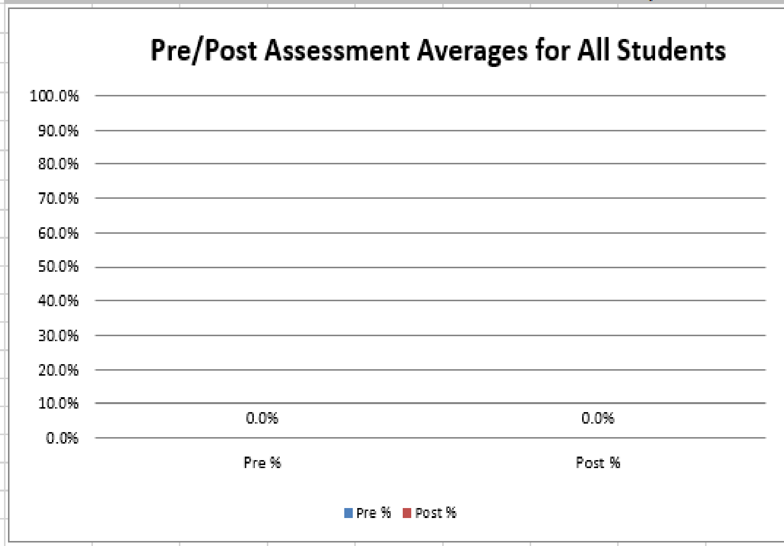


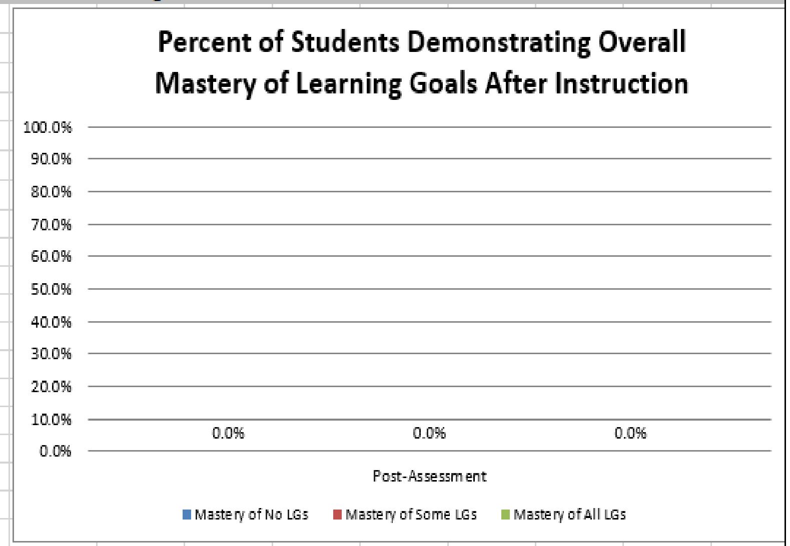
NOTE: You will want to sort your data before using this graph in your TWS. Do this by following these two steps: ***(1)*** *Return to the* ***DataEntry*** *worksheet and click in the* ***PreTotal*** *column; then select* ***Sort Smallest to Largest...*** ***(2)*** *Then, click in the* ***PostTotal*** *column and select* ***Sort Smallest to Largest...*** *again. When you return to this tab, the data will be properly sorted in your graph.*

Directions:

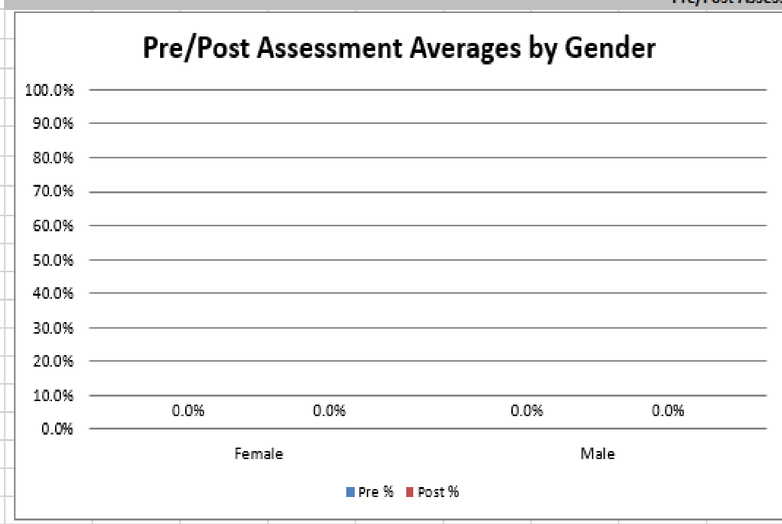
*Charts showing average pre/post assessment scores and percent of students demonstrating mastery for various student subgroups will be automatically generated here. **DO NOT** make changes on this worksheet. Instead, copy and paste the chart into Word, then change colors, labels, or titles. You may need to scroll down to find all charts.*

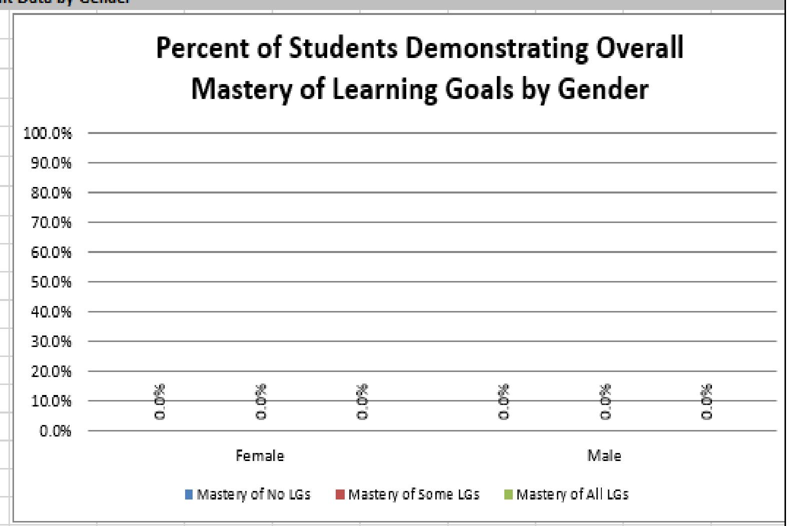
Pre/Post Assessment Data for All Students Taught



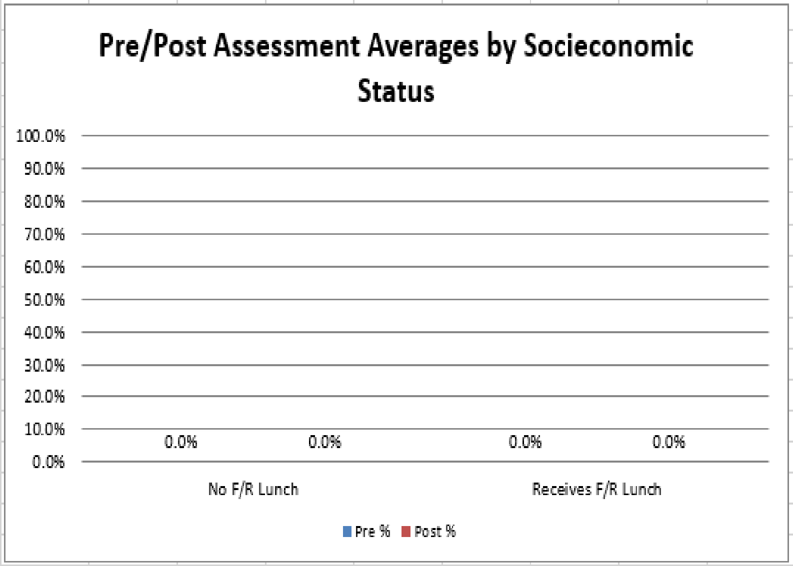


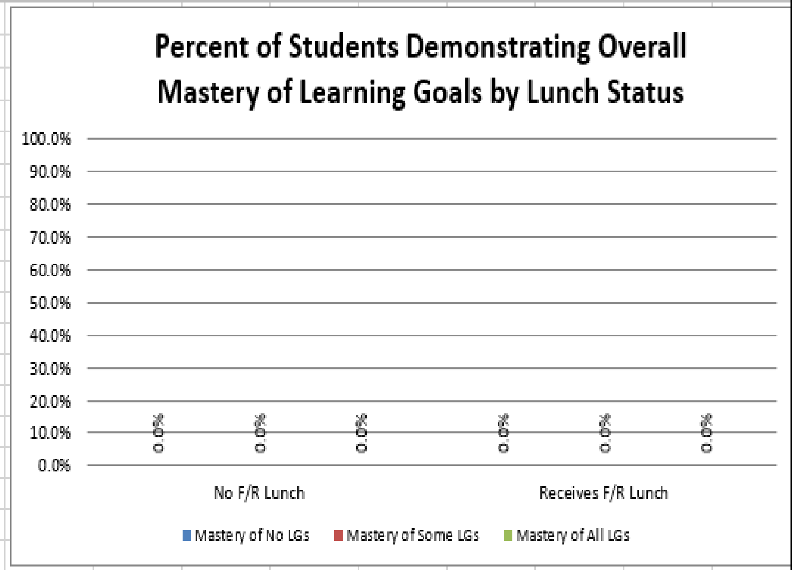
Pre/Post Assessment Data by Gender



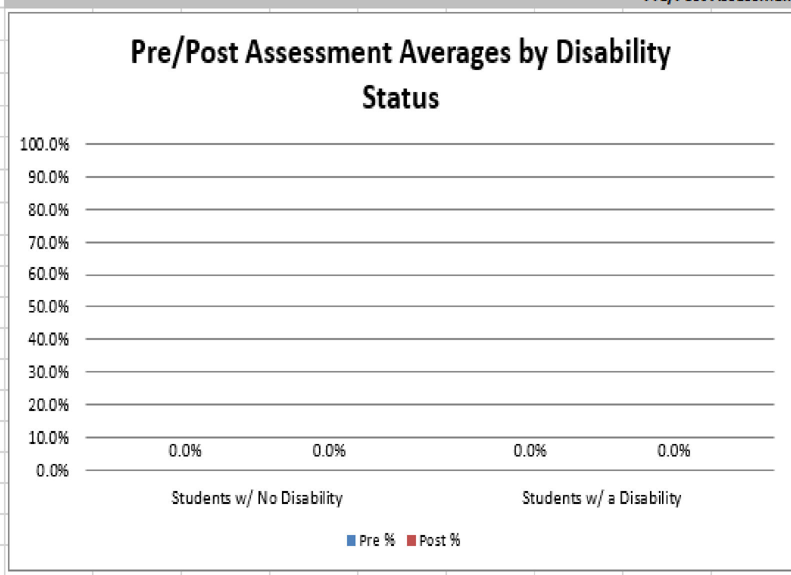


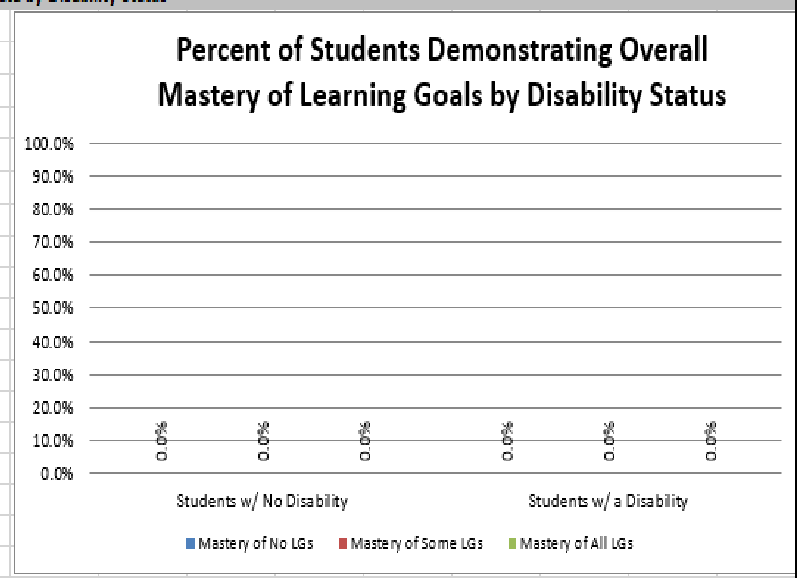
Pre/Post Assessment Data by Socieconomic Status

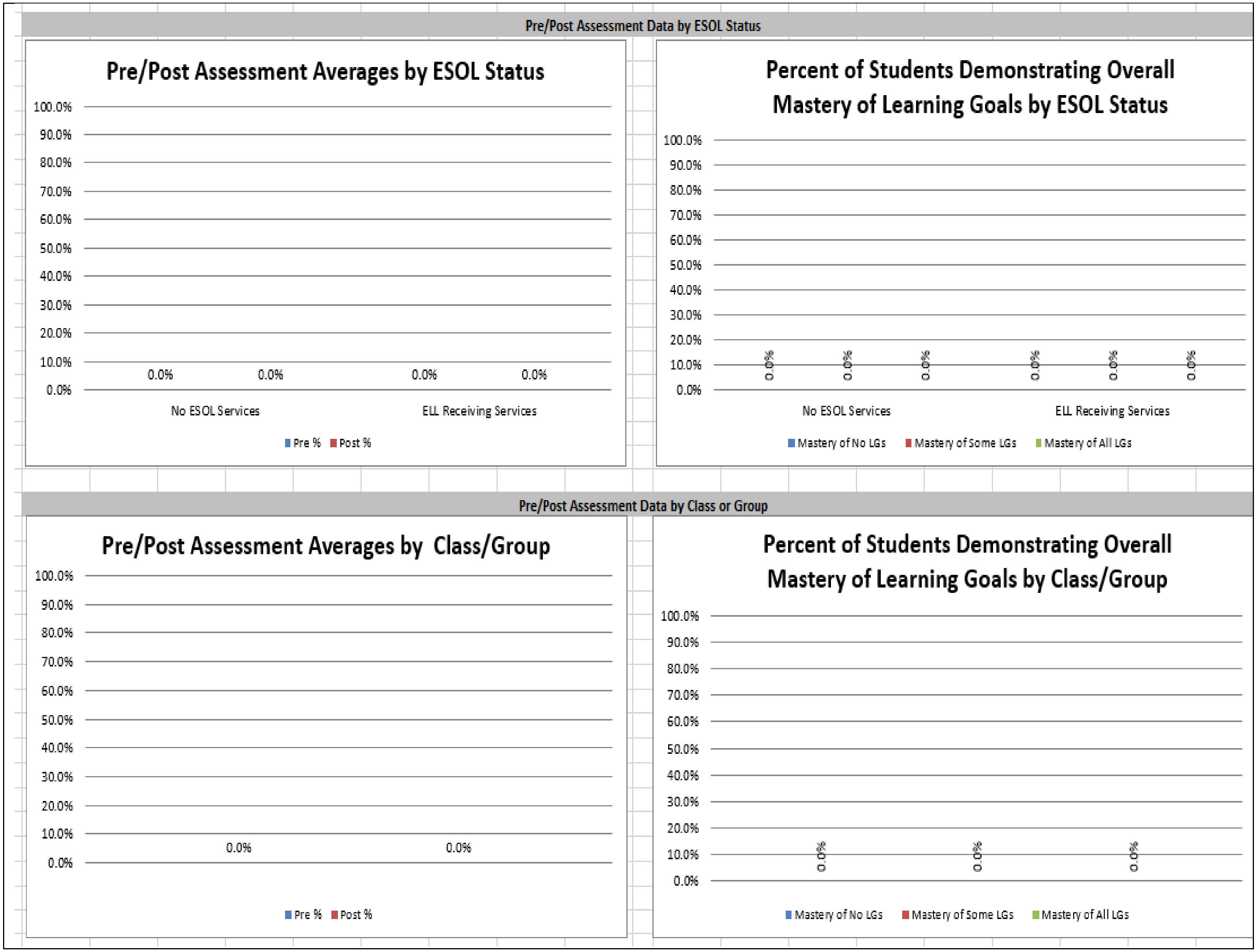




Pre/Post Assessment Data by Disability Status



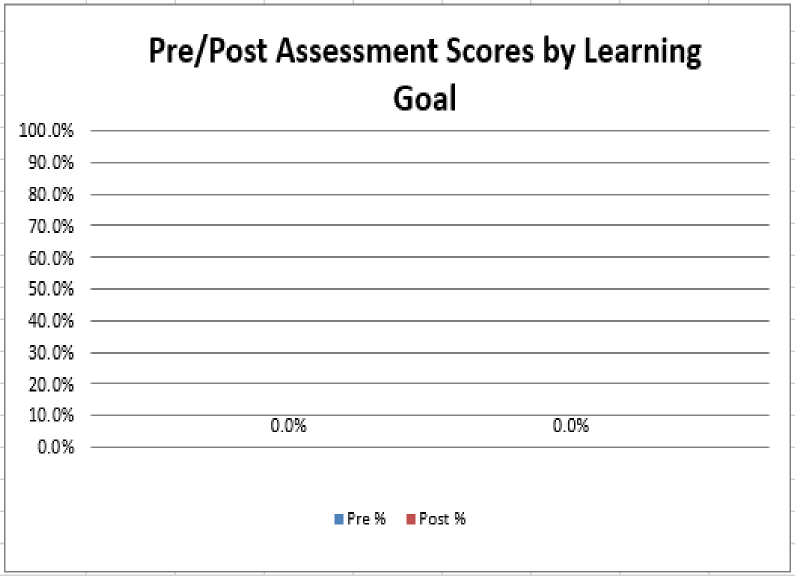


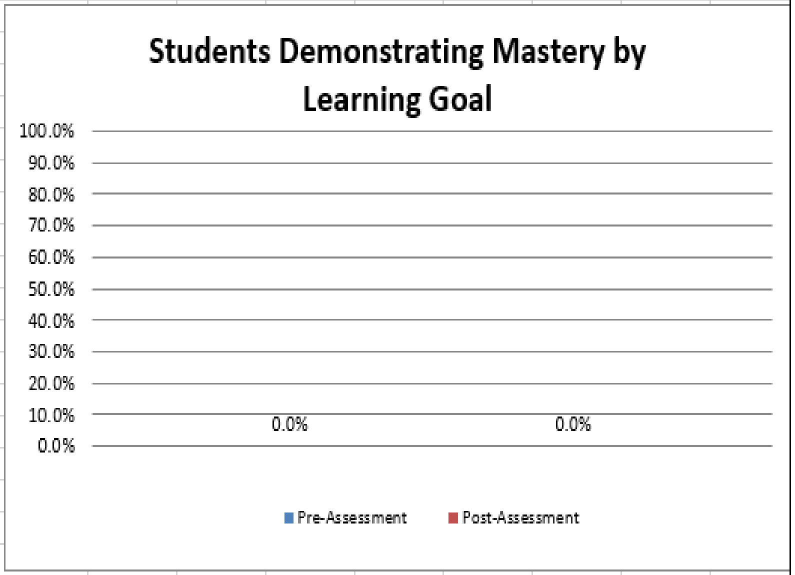

Pre/Post Assessment Data by ESOL Status
Pre/Post Assessment Averages by ESOL Status
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
0.0%
0.0%
0.0%
0.0%
No ESOL Services
ELL Receiving Services
Pre %
Post %
Percent of Students Demonstrating Overall Mastery of Learning Goals by ESOL Status
0.0%
0.0%
0.0%
0.0%
0.0%
0.0%
No ESOL Services
ELL Receiving Services
Mastery of No LGs
Mastery of Some LGs
Mastery of All LGs
Pre/Post Assessment Data by Class or Group
Pre/Post Assessment Averages by Class/Group
0.0%
0.0%
Pre %
Post %
Percent of Students Demonstrating Overall Mastery of Learning Goals by Class/Group
0.0%
0.0%
0.0%
Mastery of No LGs
Mastery of Some LGs
Mastery of All LGs

Directions:

Charts showing the average pre/post assessment scores and the percent of students who demonstrated mastery of each learning goal will be automatically generated here. ***DO NOT*** *make changes on this worksheet. Instead, copy and paste the chart into Word, then change colors, labels, or titles.*

Pre/Post Assessment Graphs by Learning Goal





APPENDIX C
SEMESTER WISE DESCRIPTIVE ANALYSES

DESCRIPTIVE ANALYSIS FOR FALL 2016

Description	Variable	N	Mean (SD) PRE-SCR	Mean(SD) POST-SCR
Independent Variables				
Student (Level 1) Variables	***STD_ID***	**3821**		
English Language Status	*EL*			
English Learners	*Yes*	270	33.54(20.32)	73.05(20.87)
Non-English Learners	*No*	3346	41.68(22.46)	79.12(17.81)
Students who exited the ESOL program	*EXIT*	205	30.10(19.08)	75.71(18.41)
Free or Reduced-Price Lunch (FRPL) Status	*FRPL*			
FRPL eligible students	*Yes*	1624	37.06(20.96)	76.07(19.28)
FRPL non-eligible students	*No*	2197	43.02(23.04)	80.31(17.05)
Ethnicity	*ETHNICITY*			
Non-Hispanic White	*W*	1745	43.02(22.92)	79.69(17.21)
Non-Hispanic Black	*B*	526	39.09(20.57)	75.66(18.73)
Hispanic, Latino or Latina	*H*	1117	35.86(21.16)	76.51(19.12)
Asian American or Pacific Islanders	*A*	176	47.54(23.34)	82.93(17.47)
American Indians or Alaskan Natives	*AI*	15	61.33(29.26)	89.08(15.35)
Other, Multiracial	*O*	139	37.42(22.37)	78.75(18.09)
Disability Status	*ESE*			
Students with exceptionalities	*Yes*	333	33.19(18.55)	72.76(20.19)
Students without exceptionalities	*No*	3252	40.68(22.25)	83.32(13.51)
Gifted Students	*GIFT*	236	48.14(25.80)	83.32(13.51)
Gender	*MALE*			
Male Students	*Yes*	1918	40.66(22.62)	78.22(17.97)
Female Students	*No*	1903	40.13(23.04)	80.31(17.05)

Description	Variable	N	Mean (SD) PRE-SCR	Mean(SD) POST-SCR
PST (Level 2) Variables	***PST_ID***			
Major	*MAJOR*			
Elementary Education	*ELEM*	2023	44.61(23.94)	79.94(18.63)
English Language Arts Education	*ELAE*	433	43.57(18.74)	81.85(16.64)
Math Education	*MATH*	341	33.11(22.04)	72.46(20.69)
Science Education	*SCIE*	582	29.67(15.62)	76.90(17.18)
Social Studies Education	*SSED*	442	38.52(19.53)	75.48(14.54)
Subject	*SUBJECT*			
Language Arts	*LA*	882	48.14(21.25)	80.54(17.40)
Math	*MA*	1199	38.56(24.62)	77.04(20.05)
Science	*SC*	1113	38.05(21.36)	79.42(17.50)
Social Studies	*SS*	627	37.74(18.49)	76.83(16.05)
Class Size	*CL_SIZE*			
Small		1073	46.65(23.57)	81.91(17.87)
Medium		787	38.51(23.50)	75.43(20.30)
Large		1961	37.91(20.53)	78.05(17.12)
Grade Level of Teaching	*GRADE_T*			
Elementary Grades	*ELEM*	2023	44.61(23.94)	79.94(18.63)
Middle School Grades	*MID*	416	33.39(20.10)	72.08(19.65)
High School Grades	*HI*	1382	36.58(19.21)	78.79(18.33)

DESCRIPTIVE ANALYSIS FOR SPRING 2017

Variable Information				
Independent Variables		Students N	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Student (Level 1) Variables	***STD_ID***	7603		
English Language Status	*EL*			
English Learners	*Yes*	472	34.30(21.62)	70.25(23.33)
Non-English Learners	*No*	6607	45.65(23.83)	79.84(18.43)
Students who exited the ESOL program	*EXIT*	246	35.64(21.19)	74.20(19.20)
Free or Reduced-Price Lunch (FRPL) Status	*FRPL*			
FRPL eligible students	*Yes*	3102	40.70(23.10)	75.87(20.48)
FRPL non-eligible students	*No*	4224	47.44(23.96)	81.36(17.43)
Ethnicity	*ETHNICITY*			
Non-Hispanic White	*W*	3282	46.66(23.99)	81.89(16.75)
Non-Hispanic Black	*B*	1231	41.69(22.89)	74.60(20.92)
Hispanic, Latino or Latina	*H*	2052	42.29(23.79)	76.59(20.17)
Asian American or Pacific Islanders	*A*	341	46.38(24.35)	81.94(17.74)
American Indians or Alaskan Natives	*AI*	40	48.10(22.55)	84.13(14.81)
Other, Multiracial	*O*	302	46.61(24.23)	80.02(19.88)
Disability Status				
Students with exceptionalities	*ESE* *Yes*	760	33.32(22.27)	73.30(22.51)
Students without exceptionalities	*No*	6014	44.92(23.55)	78.98(18.71)
Gifted Students	*GIFT*	552	56.46(22.31)	87.51(12.27)
Gender	*MALE*			
Male Students	*Yes*	3657	43.73(23.10)	78.22(19.68)
Female Students	*No*	3668	45.59(23.83)	79.84(18.21)

PST (Level 2) Variables	***PST_ID***	Students (N)	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Major	*MAJOR*			
Elementary Education	*ELEM*	*3782*	46.79(25.14)	80.72(19.19)
English Language Arts Education	*ELAE*	*1345*	46.59(22.31)	79.88(16.44)
Math Education	*MATH*	*376*	35.60(18.20)	72.03(21.19)
Science Education	*SCIE*	*507*	45.10(22.61)	76.12(18.52)
Social Studies Education	*SSED*	*1206*	37.93(21.19)	76.15(19.42)
Art Education	*ARTE*	*110*	45.46(25.05)	79.74(19.10)
Subject	*SUBJECT*			
Language Arts	*LA*	*1981*	48.83(22.79)	79.07(17.58)
Math	*MA*	*2075*	42.44(24.23)	78.84(20.08)
Science	*SC*	*1596*	49.32(24.67)	80.82(18.35)
Social Studies	*SS*	*1563*	37.15(21.26)	77.36(19.63)
Visual Arts	*VA*	*110*	45.46(25.05)	79.74(19.10)
Class Size	*CL_SIZE*			
Small		*1980*	48.24(23.91)	81.49(19.18)
Medium		*1252*	47.25(24.34)	80.93(19.18)
Large		*4093*	42.00(23.32)	77.27(18.92)
Grade Level of Teaching	*GRADE_T*			
Elementary Grades	*ELEM*	*4092*	46.71(24.82)	80.65(18.94)
Middle School Grades	*MID*	*785*	42.05(24.44)	81.16(17.95)
High School Grades	*HI*	*2449*	41.85(21.48)	75.65(18.89)

DESCRIPTIVE ANALYSIS FOR FALL 2017

Variable Information				
Independent Variables		Students *N*	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Student (Level 1) Variables	***STD_ID***	2773		
English Language Status	*EL*			
English Learners	*Yes*	298	43.53(23.09)	72.81(22.09)
Non-English Learners	*No*	2419	47.55(24.50)	77.76(18.60)
Students who exited the ESOL program	*EXIT*	56	41.91(22.27)	78.72(19.94)
Free or Reduced-Price Lunch (FRPL) Status	*FRPL*			
FRPL eligible students	*Yes*	1176	45.64(24.17)	75.50(19.10)
FRPL non-eligible students	*No*	1597	48.00(24.43)	78.53(18.98)
Ethnicity	*ETHNICITY*			
Non-Hispanic White	*W*	1283	48.96(24.09)	78.66(18.31)
Non-Hispanic Black	*B*	399	45.60(24.78)	76.05(19.14)
Hispanic, Latino or Latina	*H*	858	44.12(24.29)	74.42(20.23)
Asian American or Pacific Islanders	*A*	134	49.39(24.89)	83.60(15.31)
American Indians or Alaskan Natives	*AI*	3	63.24(18.95)	80.37(9.79)
Other, Multiracial	*O*	88	48.19(21.80)	78.34(19.00)
Disability Status	*ESE*			
Students with exceptionalities	*Yes*	297	40.60(21.06)	68.92(19.37)
Students without exceptionalities	*No*	2361	47.26(24.59)	77.66(18.93)
Gifted Students	*GIFT*	115	58.22(22.50)	90.12(10.46)
Gender	*MALE*			
Male Students	*Yes*	1401	46.81(24.45)	76.76(19.43)
Female Students	*No*	1372	47.19(24.24)	77.73(18.81)

PST (Level 2) Variables	***PST_ID***	Students	Mean (SD)	Mean (SD)
Major	*MAJOR*	*N*	PRE_SCR	POST_SCR
Elementary Education	*ELEM*	1590	48.59(24.88)	78.62(19.70)
English Language Arts Education	*ELAE*	372	52.26(20.40)	70.23(18.40)
Math Education	*MATH*	264	29.38(20.57)	78.38(16.06)
Science Education	*SCIE*	151	46.84(22.24)	72.85(17.57)
Social Studies Education	*SSED*	201	35.59(19.95)	75.24(19.48)
Art Education	*ARTE*	162	62.42(20.24)	81.97(14.88)
Foreign Language Education	*FLDE*	33	46.82(19.87)	90.33(14.55)
Subject	*SUBJECT*			
Language Arts	*LA*	758	51.17(22.19)	71.43(20.15)
Math	*MA*	908	43.76(25.74)	79.91(19.04)
Science	*SC*	445	51.76(23.22)	78.15(17.77)
Social Studies	*SS*	467	36.66(22.19)	78.07(17.84)
Visual Arts	*VA*	162	62.42(20.24)	81.97(14.88)
World Language	*WL*	33	46.82(19.88)	90.33(14.55)
Class Size	*CL_SIZE*			
Small		717	53.86(25.01)	79.97(19.70)
Medium		718	43.71(24.28)	78.07(19.80)
Large		1338	45.09(23.29)	75.34(18.14)
Grade Level of Teaching	*GRADE_T*			
Elementary Grades	*ELEM*	1555	48.80(24.85)	77.83(19.91)
Middle School Grades	*MID*	293	48.45(21.56)	78.44(15.83)
High School Grades	*HI*	925	43.51(23.95)	75.88(18.55)

DESCRIPTIVE ANALYSIS FOR SPRING 2018

Variable Information				
Independent Variables		Students *N*	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Student (Level 1) Variables	***STD_ID***	4807		
English Language Status	*EL*			
English Learners	*Yes*	404	44.13(24.99)	74.72(21.31)
Non-English Learners	*No*	4283	50.36(24.08)	82.65(17.46)
Students who exited the ESOL program	*EXIT*	120	47.62(23.57)	79.80(15.97)
Free or Reduced-Price Lunch (FRPL) Status	*FRPL*			
FRPL eligible students	*Yes*	2395	49.12(23.81)	80.40(19.34)
FRPL non-eligible students	*No*	2412	50.42(24.58)	83.41(12.25)
Ethnicity	*ETHNICITY*			
Non-Hispanic White	*W*	1899	50.59(24.13)	84.58(15.95)
Non-Hispanic Black	*B*	867	47.41(24.77)	78.97(20.12)
Hispanic, Latino or Latina	*H*	1435	47.43(24.02)	80.11(18.52)
Asian American or Pacific Islanders	*A*	180	49.61(25.50)	84.08(16.57)
American Indians or Alaskan Natives	*AI*	11	50.24(25.73)	80.91(12.86)
Other, Multiracial	*O*	195	58.25(23.36)	84.90(15.31)
Disability Status	*ESE*			
Students with exceptionalities	*Yes*	471	43.94(24.02)	75.00(20.54)
Students without exceptionalities	*No*	4022	49.60(24.00)	82.18(17.60)
Gifted Students	*GIFT*	314	60.65(23.71)	88.82(14.01)
Gender	*MALE*			
Male Students	*Yes*	2376	49.70(24.62)	81.98(17.88)
Female Students	*No*	2345	49.51(23.95)	82.36(17.78)

PST (Level 2) Variables	***PST_ID***	Student *N*	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Major	*MAJOR*			
Elementary Education	*ELEM*	3936	50.77(24.91)	83.03(17.58)
Social Studies Education	*SSED*	871	45.22(20.10)	76.87(18.55)
Subject	*SUBJECT*			
Language Arts	*LA*	981	55.50(26.68)	83.58(17.71)
Math	*MA*	1616	49.14(24.45)	81.96(18.15)
Science	*SC*	912	52.90(23.51)	84.90(16.09)
Social Studies	*SS*	1278	43.49(20.41)	78.61(18.22)
Class Size	*CL_SIZE*			
Small		2059	51.20(24.00)	83.33(18.26)
Medium		1163	53.73(23.77)	82.76(17.74)
Large		1585	44.99(24.04)	79.45(17.34)
Grade Level of Teaching	*GRADE_T*			
Elementary Grades	*ELEM*	3754	51.33(25.00)	83.20(17.67)
Middle School Grades	*MID*	378	33.76(19.61)	78.23(15.67)
High School Grades	*HI*	675	50.01(18.04)	76.80(19.24)

DESCRIPTIVE ANALYSIS FOR FALL 2018

Variable Information				
Independent Variables		Student *N*	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Student (Level 1) Variables	***STD_ID***	2082		
English Language Status	*EL*			
English Learners	*Yes*	290	41.28(22.16)	67.18(23.08)
Non-English Learners	*No*	1716	50.85(25.71)	79.23(20.52)
Students who exited the ESOL program	*EXIT*	76	47.70(25.10)	69.47(20.70)
Free or Reduced-Price Lunch (FRPL) Status	*FRPL*			
FRPL eligible students	*Yes*	1272	46.26(24.25)	74.57(22.01)
FRPL non-eligible students	*No*	810	54.34(26.45)	81.33(19.56)
Ethnicity	*ETHNICITY*			
Non-Hispanic White	*W*	774	52.54(26.86)	80.73(20.32)
Non-Hispanic Black	*B*	358	47.71(24.86)	75.10(21.08)
Hispanic, Latino or Latina	*H*	728	47.26(24.06)	74.12(21.81)
Asian American or Pacific Islanders	*A*	59	49.93(21.28)	78.62(22.36)
American Indians or Alaskan Natives	*AI*	12	41.82(16.85)	90.16(11.43)
Other, Multiracial	*O*	107	52.31(25.32)	71.51(22.72)
Disability Status	*ESE*			
Students with exceptionalities	*Yes*	269	45.34(24.23)	69.31(23.04)
Students without exceptionalities	*No*	1712	49.11(25.32)	77.95(20.79)
Gifted Students	*GIFT*	101	65.12(24.86)	85.37(20.68)
Gender	*MALE*			
Male Students	*Yes*	1030	48.10(25.28)	78.28(20.86)
Female Students	*No*	1052	50.68(25.52)	76.09(21.80)

PST (Level 2) Variables	***PST_ID***	Student *N*	Mean (SD) PRE_SCR	Mean (SD) POST_SCR
Major	*MAJOR*			
Elementary Education	*ELEM*	2082	49.40(25.43)	77.20(21.35)
Subject	*SUBJECT*			
Language Arts	*LA*	852	52.37(25.04)	75.26(21.92)
Math	*MA*	534	46.44(26.56)	76.35(22.76)
Science	*SC*	574	50.47(23.81)	80.64(18.05)
Social Studies	*SS*	122	36.65(25.49)	78.25(23.44)
Class Size	*CL_SIZE*			
Small		1210	49.10(24.04)	77.35(20.58)
Medium		510	51.60(26.61)	80.76(20.89)
Large		362	47.32(27.96)	71.68(23.58)

LEVEL 1, LEVEL 2, LEVEL 3, HLM DESCRIPTIVE ANALYSIS

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MALE	20703	0.50	0.50	0.00	1.00
ETHNICIT	20358	1.13	1.26	0.00	5.00
BLACK	20789	0.16	0.37	0.00	1.00
HISPANIC	20789	0.30	0.46	0.00	1.00
ASIAN	20789	0.04	0.20	0.00	1.00
AI	20789	0.00	0.06	0.00	1.00
OTHER	20789	0.04	0.20	0.00	1.00
FRPL	20789	0.46	0.50	0.00	1.00
ESE	20789	0.23	0.55	0.00	2.00
DISABLE	20789	0.10	0.30	0.00	1.00
GIFT	20789	0.06	0.24	0.00	1.00
EL	20789	0.15	0.44	0.00	2.00
ELS	20789	0.08	0.28	0.00	1.00
EXIT	20789	0.03	0.18	0.00	1.00
PRE_SCR	20789	45.80	24.09	0.00	100.00

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MAJOR	767	0.47	1.20	0.00	6.00
ELAE	767	0.04	0.20	0.00	1.00
MATH	767	0.03	0.17	0.00	1.00
SCIE	767	0.02	0.14	0.00	1.00
SSED	767	0.06	0.23	0.00	1.00
ARTE	767	0.01	0.12	0.00	1.00
FLDE	767	0.00	0.04	0.00	1.00
SUBJECT	767	1.29	1.02	0.00	5.00
MA	767	0.38	0.49	0.00	1.00
SC	767	0.23	0.42	0.00	1.00
SS	767	0.13	0.33	0.00	1.00
VA	767	0.01	0.12	0.00	1.00
WL	767	0.00	0.04	0.00	1.00
GRADE	767	0.27	0.65	0.00	2.00
MIDDLE	767	0.05	0.21	0.00	1.00
HIGH	767	0.11	0.31	0.00	1.00
CL_SIZE	767	0.66	0.80	0.00	2.00
MEDIUM	767	0.26	0.44	0.00	1.00
LARGE	767	0.20	0.40	0.00	1.00

LEVEL-3 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
S2017	5	0.20	0.45	0.00	1.00
F2017	5	0.20	0.45	0.00	1.00
S2018	5	0.20	0.45	0.00	1.00
F2018	5	0.20	0.45	0.00	1.00

APPENDIX D
HLM OUTPUTS

HLM OUTPUT FOR FINAL POSTTEST MODEL

Program:	HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors:	Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:	Scientific Software International, Inc. (c) 2013
	hlm@ssicentral.com
	www.ssicentral.com

Module:	HLM3.EXE (7.03.21710.17001)
Date:	22 May 2020, Friday
Time:	23: 6:53

Specifications for this HLM3 run

Problem Title: no title

The data source for this run = Posttest.mdmt
The command file for this run = \\net.ucf.edu\Users\KnightsDrive\ni728415\Desktop\Differential Analysis\Posttest\Postfinal.hlm
Output file name = \\net.ucf.edu\Users\KnightsDrive\ni728415\Desktop\Differential Analysis\Posttest\hlm3.html
The maximum number of level-1 units = 20789
The maximum number of level-2 units = 767
The maximum number of level-3 units = 5
The maximum number of iterations = 100
Method of estimation: full maximum likelihood
The outcome variable is POST_SCR

Summary of the model specified

Level-1 Model

$$\mathrm{POST\_SCR}_{mij} = \psi_{0ij} + \psi_{1ij}*(BLACK_{mij}) + \psi_{2ij}*(HISPANIC_{mij}) + \psi_{3ij}*(ASIAN_{mij}) + \psi_{4ij}*(FRPL_{mij}) + \psi_{5ij}*(DISABLE_{mij}) + \psi_{6ij}*(GIFT_{mij}) + \psi_{7ij}*(ELS_{mij}) + \psi_{8ij}*(EXIT_{mij}) + \psi_{9ij}*(PRE\_SCR_{mij}) + \varepsilon_{mij}$$

Level-2 Model

$$\psi_{0ij} = \pi_{00j} + \pi_{01j}*(SUBJECT_{ij}) + \pi_{02j}*(CL\_SIZE_{ij}) + e_{0ij}$$
$$\psi_{1ij} = \pi_{10j} + e_{1ij}$$
$$\psi_{2ij} = \pi_{20j} + e_{2ij}$$

$\psi_{3ij} = \pi_{30j}$
$\psi_{4ij} = \pi_{40j} + \pi_{41j}*(GRADE_{ij}) + e_{4ij}$
$\psi_{5ij} = \pi_{50j} + \pi_{51j}*(SUBJECT_{ij}) + \pi_{52j}*(CL\_SIZE_{ij})$
$\psi_{6ij} = \pi_{60j} + \pi_{61j}*(GRADE_{ij})$
$\psi_{7ij} = \pi_{70j} + e_{7ij}$
$\psi_{8ij} = \pi_{80j}$
$\psi_{9ij} = \pi_{90j}$

Level-3 Model

$\pi_{00j} = \beta_{000} + r_{00j}$
$\pi_{01j} = \beta_{010}$
$\pi_{02j} = \beta_{020}$
$\pi_{10j} = \beta_{100}$
$\pi_{20j} = \beta_{200}$
$\pi_{30j} = \beta_{300}$
$\pi_{40j} = \beta_{400}$
$\pi_{41j} = \beta_{410}$
$\pi_{50j} = \beta_{500}$
$\pi_{51j} = \beta_{510}$
$\pi_{52j} = \beta_{520}$
$\pi_{60j} = \beta_{600}$
$\pi_{61j} = \beta_{610}$
$\pi_{70j} = \beta_{700} + r_{70j}$
$\pi_{80j} = \beta_{800}$
$\pi_{90j} = \beta_{900}$

BLACK HISPANIC ASIAN FRPL DISABLE GIFT ELS EXIT have been centered around the group mean.

PRE_SCR has been centered around the grand mean.

SUBJECT GRADE CL_SIZE have been centered around the group mean.

Mixed Model

$POST\_SCR_{mij} = \beta_{000} + \beta_{010}*\mathrm{SUBJECT}_{ij} + \beta_{020}*\mathrm{CL\_SIZE}_{ij} + \beta_{100}*BLACK_{mij}$
$+ \beta_{200}*HISPANIC_{mij} + \beta_{300}*ASIAN_{mij} + \beta_{400}*FRPL_{mij} + \beta_{410}*FRPL_{mij}*\mathrm{GRADE}_{ij}$
$+ \beta_{500}*DISABLE_{mij} + \beta_{510}*DISABLE_{mij}*\mathrm{SUBJECT}_{ij} + \beta_{520}*DISABLE_{mij}*\mathrm{CL\_SIZE}_{ij} + \beta_{600}*GIFT_{mij}$
$+ \beta_{610}*GIFT_{mij}*\mathrm{GRADE}_{ij} + \beta_{700}*ELS_{mij} + \beta_{800}*EXIT_{mij} + \beta_{900}*PRE\_SCR_{mij}$
$+ e_{0ij} + e_{1ij} *BLACK_{mij} + e_{2ij} *HISPANIC_{mij} + e_{4ij}$
$*FRPL_{mij} + e_{7ij} *ELS_{mij}$
$+ r_{00j} + r_{70j} *ELS_{mij} + \varepsilon_{mij}$

Run-time deletion has reduced the number of level-1 records to 20788

For starting values, data from 6361 level-1 and 216 level-2 records were used

Final Results - Iteration 73

Iterations stopped due to small change in likelihood function

σ^2_ε = 201.71160

Standard error of σ^2_ε = 2.08805

σ^2_e

INTRCPT1,ψ_0	90.66241	5.80171	5.38933	3.83856	21.05282
BLACK,ψ_1	5.80171	9.09755	1.33777	2.90191	-0.08857
HISPANIC,ψ_2	5.38933	1.33777	1.11098	0.89037	-4.15724
FRPL,ψ_4	3.83856	2.90191	0.89037	3.65062	2.05077
ELS,ψ_7	21.05282	-0.08857	-4.15724	2.05077	59.50367

Standard errors of σ^2_e

INTRCPT1,ψ_0	5.13340	3.58918	2.73274	3.21906	5.94398
BLACK,ψ_1	3.58918	3.92301	2.23446	2.53148	5.61320
HISPANIC,ψ_2	2.73274	2.23446	2.07019	1.80050	4.22470
FRPL,ψ_4	3.21906	2.53148	1.80050	2.47509	5.01056
ELS,ψ_7	5.94398	5.61320	4.22470	5.01056	10.10878

σ^2_e (as correlations)

INTRCPT1,ψ_0	1.000	0.202	0.537	0.211	0.287
BLACK,ψ_1	0.202	1.000	0.421	0.504	-0.004
HISPANIC,ψ_2	0.537	0.421	1.000	0.442	-0.511
FRPL,ψ_4	0.211	0.504	0.442	1.000	0.139
ELS,ψ_7	0.287	-0.004	-0.511	0.139	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1,ψ_0	0.910
BLACK,ψ_1	0.099
HISPANIC,ψ_2	0.021
FRPL,ψ_4	0.072
ELS,ψ_7	0.327

Note: The reliability estimates reported above are based on only 216 of 767 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

τ_π

INTRCPT1	ELS
INTRCPT2, π_{00}	INTRCPT2, π_{70}
2.20506	0.54955
0.54955	0.13749

Standard errors of τ_π

INTRCPT1	ELS
INTRCPT2, π_{00}	INTRCPT2, π_{70}
1.84027	1.05726
1.05726	1.08522

τ_π (as correlations)

INTRCPT1/INTRCPT2, π_{00}	1.000	0.998
ELS/INTRCPT2, π_{70}	0.998	1.000

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, π_{00}	0.755
ELS/INTRCPT2, π_{70}	0.077

The value of the log-likelihood function at iteration 73 = -8.576298E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	79.139712	0.765107	103.436	4	<0.001
For SUBJECT, π_{01}					
INTRCPT3, β_{010}	1.501517	0.352376	4.261	751	<0.001
For CL_SIZE, π_{02}					
INTRCPT3, β_{020}	-0.981347	0.451967	-2.171	751	0.030
For BLACK slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	-1.690760	0.364779	-4.635	751	<0.001
For HISPANIC slope, ψ_2					
For INTRCPT2, π_{20}					
INTRCPT3, β_{200}	-1.047721	0.285194	-3.674	751	<0.001
For ASIAN slope, ψ_3					
For INTRCPT2, π_{30}					
INTRCPT3, β_{300}	0.920609	0.522552	1.762	16935	0.078
For FRPL slope, ψ_4					
For INTRCPT2, π_{40}					
INTRCPT3, β_{400}	-2.207513	0.354860	-6.221	751	<0.001

For GRADE, π_{41}					
INTRCPT3, β_{410}	0.440175	0.374707	1.175	751	0.240
For DISABLE slope, ψ_5					
For INTRCPT2, π_{50}					
INTRCPT3, β_{500}	-5.650378	0.406596	-13.897	16935	<0.001
For SUBJECT, π_{51}					
INTRCPT3, β_{510}	-0.154219	0.331916	-0.465	16935	0.642
For CL_SIZE, π_{52}					
INTRCPT3, β_{520}	1.385807	0.432742	3.202	16935	0.001
For GIFT slope, ψ_6					
For INTRCPT2, π_{60}					
INTRCPT3, β_{600}	3.228388	0.532606	6.061	16935	<0.001
For GRADE, π_{61}					
INTRCPT3, β_{610}	-0.571703	0.610553	-0.936	16935	0.349
For ELS slope, ψ_7					
For INTRCPT2, π_{70}					
INTRCPT3, β_{700}	-4.396846	0.617001	-7.126	4	0.002
For EXIT slope, ψ_8					
For INTRCPT2, π_{80}					
INTRCPT3, β_{800}	-0.366777	0.615055	-0.596	16935	0.551
For PRE_SCR slope, ψ_9					
For INTRCPT2, π_{90}					
INTRCPT3, β_{900}	0.307293	0.005604	54.835	16935	<0.001

The robust standard errors cannot be computed for this model.

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, e_0	9.52168	90.66241	209	2631.21303	<0.001
BLACK slope, e_1	3.01621	9.09755	215	250.72970	0.048
HISPANIC slope, e_2	1.05403	1.11098	215	221.80790	0.360
FRPL slope, e_4	1.91066	3.65062	214	241.16176	0.098
ELS slope, e_7	7.71386	59.50367	211	413.79501	<0.001
level-1, ε	14.20252	201.71160			

Note: The chi-square statistics reported above are based on only 216 of 767 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of level-3 variance components

Random Effect	*d.f.*	χ^2	*p*-value

	Standard Deviation	Variance Component			
INTRCPT1/INTRCPT2,r_{00}	1.48494	2.20506	4	21.07637	<0.001
ELS/INTRCPT2,r_{70}	0.37080	0.13749	4	4.76464	0.312

Statistics for the current model

Deviance = 171525.966158
Number of estimated parameters = 35

HLM OUTPUT FOR LONGITUDINAL STUDY

Program:	HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors:	Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:	Scientific Software International, Inc. (c) 2013
	hlm@ssicentral.com
	www.ssicentral.com

Module:	HLM2.EXE (7.03.21710.17001)
Date:	23 May 2020, Saturday
Time:	21:38: 8

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = LTNull.mdmt
The command file for this run = C:\Users\ni728415\AppData\Local\Temp\whlmtemp.hlm
Output file name = \\net.ucf.edu\Users\KnightsDrive\ni728415\Desktop\Differential Analysis\Longterm Trend\hlm2.html
The maximum number of level-1 units = 20376
The maximum number of level-2 units = 749
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is POST_SCR

Summary of the model specified

Level-1 Model

$$POST\_SCR_{ti} = \pi_{0i} + \pi_{1i}*(TIM\_E\_ID_{ti}) + \pi_{2i}*(PRE\_SCR_{ti}) + e_{ti}$$

Level-2 Model

$$\begin{aligned}\pi_{0i} = \beta_{00} &+ \beta_{01}*(MA_i) + \beta_{02}*(SC_i) + \beta_{03}*(MIDDLE_i) + \beta_{04}*(HIGH_i) \\ &+ \beta_{05}*(MEDIUM_i) + \beta_{06}*(LARGE_i) + \beta_{07}*(BLACK_i) + \beta_{08}*(HISPANIC_i) \\ &+ \beta_{09}*(ASIAN_i) + \beta_{010}*(AI_i) + \beta_{011}*(OTHER_i) + \beta_{012}*(FRPL_i) \\ &+ \beta_{013}*(DISABLE_i) + \beta_{014}*(GIFT_i) + \beta_{015}*(EL_i) + r_{0i}\end{aligned}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}*(MA_i) + \beta_{12}*(SC_i) + \beta_{13}*(MIDDLE_i) + \beta_{14}*(HIGH_i)$$

$+ \beta_{15}*(MEDIUM_i) + \beta_{16}*(LARGE_i) + \beta_{17}*(BLACK_i) + \beta_{18}*(HISPANIC_i)$
$+ \beta_{19}*(ASIAN_i) + \beta_{110}*(AI_i) + \beta_{111}*(OTHER_i) + \beta_{112}*(FRPL_i)$
$+ \beta_{113}*(DISABLE_i) + \beta_{114}*(GIFT_i) + \beta_{115}*(EL_i)$
$\pi_{2i} = \beta_{20} + r_{2i}$

PRE_SCR has been centered around the group mean.

MA SC MIDDLE HIGH MEDIUM LARGE BLACK HISPANIC ASIAN AI OTHER FRPL DISABLE GIFT EL have been centered around the grand mean.

Mixed Model

$POST\_SCR_{ti} = \beta_{00} + \beta_{01}*MA_i + \beta_{02}*SC_i + \beta_{03}*MIDDLE_i$
$+ \beta_{04}*HIGH_i + \beta_{05}*MEDIUM_i + \beta_{06}*LARGE_i + \beta_{07}*BLACK_i$
$+ \beta_{08}*HISPANIC_i + \beta_{09}*ASIAN_i + \beta_{010}*AI_i + \beta_{011}*OTHER_i$
$+ \beta_{012}*FRPL_i + \beta_{013}*DISABLE_i + \beta_{014}*GIFT_i + \beta_{015}*EL_i$
$+ \beta_{10}*TIM\_E\_ID_{ti} + \beta_{11}*MA_i*TIM\_E\_ID_{ti} + \beta_{12}*SC_i*TIM\_E\_ID_{ti} + \beta_{13}*MIDDLE_i*TIM\_E\_ID_{ti}$
$+ \beta_{14}*HIGH_i*TIM\_E\_ID_{ti} + \beta_{15}*MEDIUM_i*TIM\_E\_ID_{ti} + \beta_{16}*LARGE_i*TIM\_E\_ID_{ti} +$
$\beta_{17}*BLACK_i*TIM\_E\_ID_{ti}$
$+ \beta_{18}*HISPANIC_i*TIM\_E\_ID_{ti} + \beta_{19}*ASIAN_i*TIM\_E\_ID_{ti} + \beta_{110}*AI_i*TIM\_E\_ID_{ti} +$
$\beta_{111}*OTHER_i*TIM\_E\_ID_{ti}$
$+ \beta_{112}*FRPL_i*TIM\_E\_ID_{ti} + \beta_{113}*DISABLE_i*TIM\_E\_ID_{ti} + \beta_{114}*GIFT_i*TIM\_E\_ID_{ti} +$
$\beta_{115}*EL_i*TIM\_E\_ID_{ti}$
$+ \beta_{20}*PRE\_SCR_{ti}$
$+ r_{0i} + r_{2i}*PRE\_SCR_{ti} + e_{ti}$

Run-time deletion has reduced the number of level-1 records to 20375

Final Results - Iteration 4

Iterations stopped due to small change in likelihood function

$\sigma^2 = 198.51895$

τ

INTRCPT1,π_0	106.10342	-1.05621
PRE_SCR,π_2	-1.05621	0.04308

τ (as correlations)

INTRCPT1,π_0	1.000	-0.494
PRE_SCR,π_2	-0.494	1.000

Random level-1 coefficient	Reliability estimate

INTRCPT1, π_0	0.918
PRE_SCR, π_2	0.557

The value of the log-likelihood function at iteration 4 = -8.403033E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	80.128355	0.930408	86.122	733	<0.001
MA, β_{01}	-0.835649	2.115404	-0.395	733	0.693
SC, β_{02}	1.621053	2.371881	0.683	733	0.495
MIDDLE, β_{03}	-7.215460	4.323708	-1.669	733	0.096
HIGH, β_{04}	-3.595720	3.295074	-1.091	733	0.276
MEDIUM, β_{05}	-6.610158	2.251921	-2.935	733	0.003
LARGE, β_{06}	-5.132098	2.779363	-1.847	733	0.065
BLACK, β_{07}	-1.492586	2.687937	-0.555	733	0.579
HISPANIC, β_{08}	0.104023	2.281965	0.046	733	0.964
ASIAN, β_{09}	-1.970017	5.861432	-0.336	733	0.737
AI, β_{010}	6.017590	17.734181	0.339	733	0.734
OTHER, β_{011}	3.589784	4.406195	0.815	733	0.416
FRPL, β_{012}	-2.050106	1.920431	-1.068	733	0.286
DISABLE, β_{013}	3.432586	2.321975	1.478	733	0.140
GIFT, β_{014}	8.928611	4.169123	2.142	733	0.033
EL, β_{015}	-2.201517	1.785300	-1.233	733	0.218
For TIM_E_ID slope, π_1					
INTRCPT2, β_{10}	0.102770	0.298017	0.345	18861	0.730
MA, β_{11}	0.746663	0.663338	1.126	18861	0.260
SC, β_{12}	0.604824	0.733205	0.825	18861	0.409
MIDDLE, β_{13}	2.114676	1.640806	1.289	18861	0.197
HIGH, β_{14}	1.119665	1.289870	0.868	18861	0.385
MEDIUM, β_{15}	1.862133	0.700117	2.660	18861	0.008
LARGE, β_{16}	0.671591	0.898518	0.747	18861	0.455
BLACK, β_{17}	-0.293618	0.844375	-0.348	18861	0.728
HISPANIC, β_{18}	-0.486207	0.730949	-0.665	18861	0.506
ASIAN, β_{19}	1.429447	1.911134	0.748	18861	0.454
AI, β_{110}	-3.020779	6.231337	-0.485	18861	0.628
OTHER, β_{111}	-0.955775	1.261670	-0.758	18861	0.449
FRPL, β_{112}	-0.001161	0.621644	-0.002	18861	0.999
DISABLE, β_{113}	-1.298989	0.727857	-1.785	18861	0.074
GIFT, β_{114}	-2.744958	1.213148	-2.263	18861	0.024
EL, β_{115}	0.672141	0.620530	1.083	18861	0.279
For PRE_SCR slope, π_2					

INTRCPT2, β_{20}	0.368874	0.010103	36.512	748	<0.001

The robust standard errors cannot be computed for this model.

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, r_0	10.30065	106.10342	733	11132.65326	<0.001
PRE_SCR slope, r_2	0.20757	0.04308	748	2098.00002	<0.001
level-1, e	14.08968	198.51895			

Statistics for current covariance components model

Deviance = 168060.656175
Number of estimated parameters = 4

APPENDIX E
PRELIMINARY ANALYSES

LEVEL 1 RESIDUAL ANALYSIS

Table A5.1.

Residual Analysis for Level 1 Normality.

			Statistics	Std. Error
I1resid	Mean		-.00001	.116783
	95% confidence Interval for Mean	Lower Bound	-.22892	
		Upper Bound	.22889	
	5% Trimmed Mean		-.07993	
	Median		-.22871	
	Variance		283.524	
	Std. Deviation		16.838170	
	Minimum		-85.654	
	Maximum		71.431	
	Range		157.085	
	Interquartile Range		20.683	
	Skewness		.077	.017
	Kurtosis		.763	.034

Table A5.2.

Test of Level 1 Normality.

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
l1resid	.023	20789	.000

a. Lilliefors Significance Correction

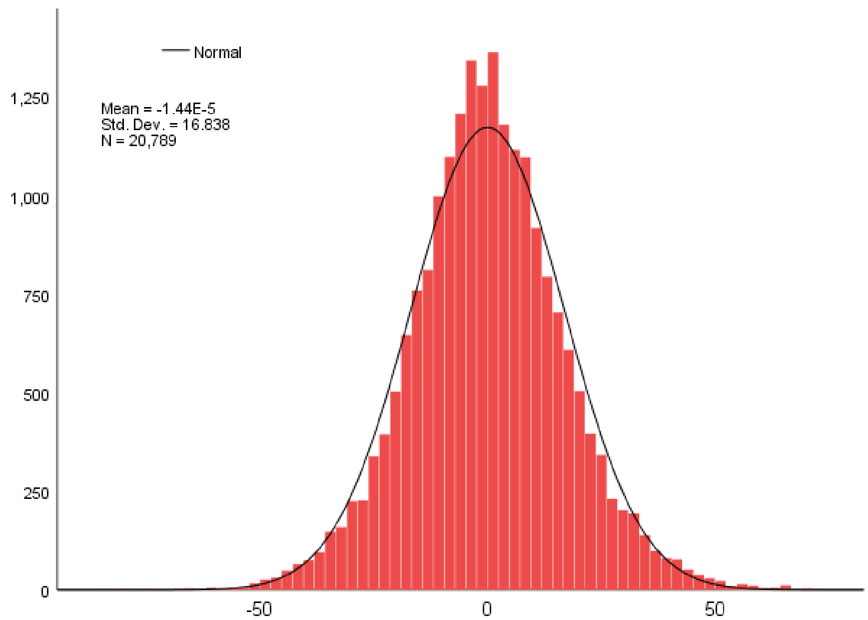


Figure A5.1a. Histogram with a normal curve for Level 1 residual showing Level 1 Normality.

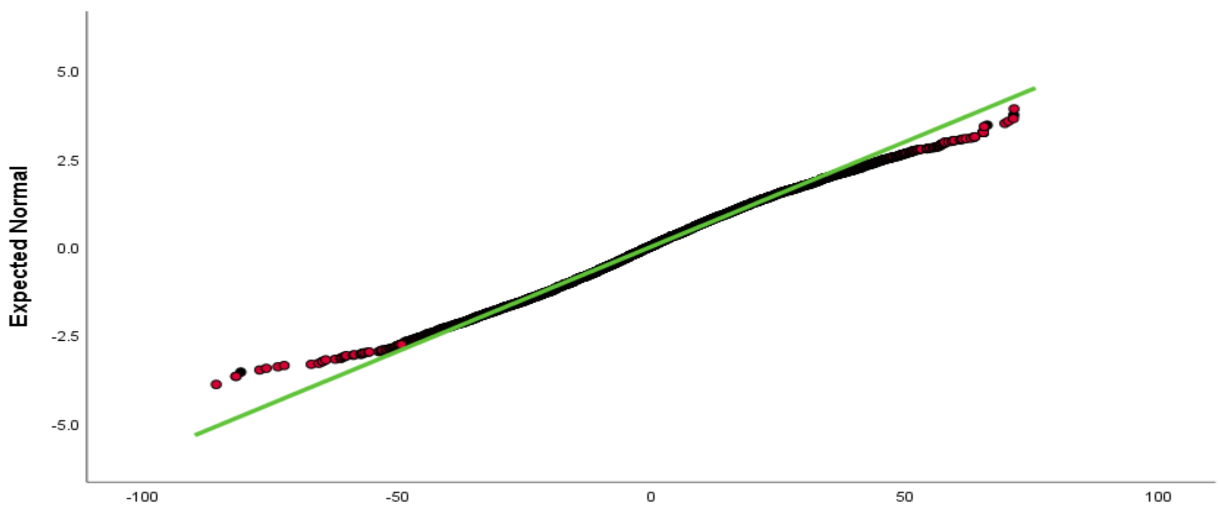


Figure A5.1b. Normal Q-Q Plot of Level 1 residual based on the discrepancies between the observed and fitted values showing Level 1 normality.

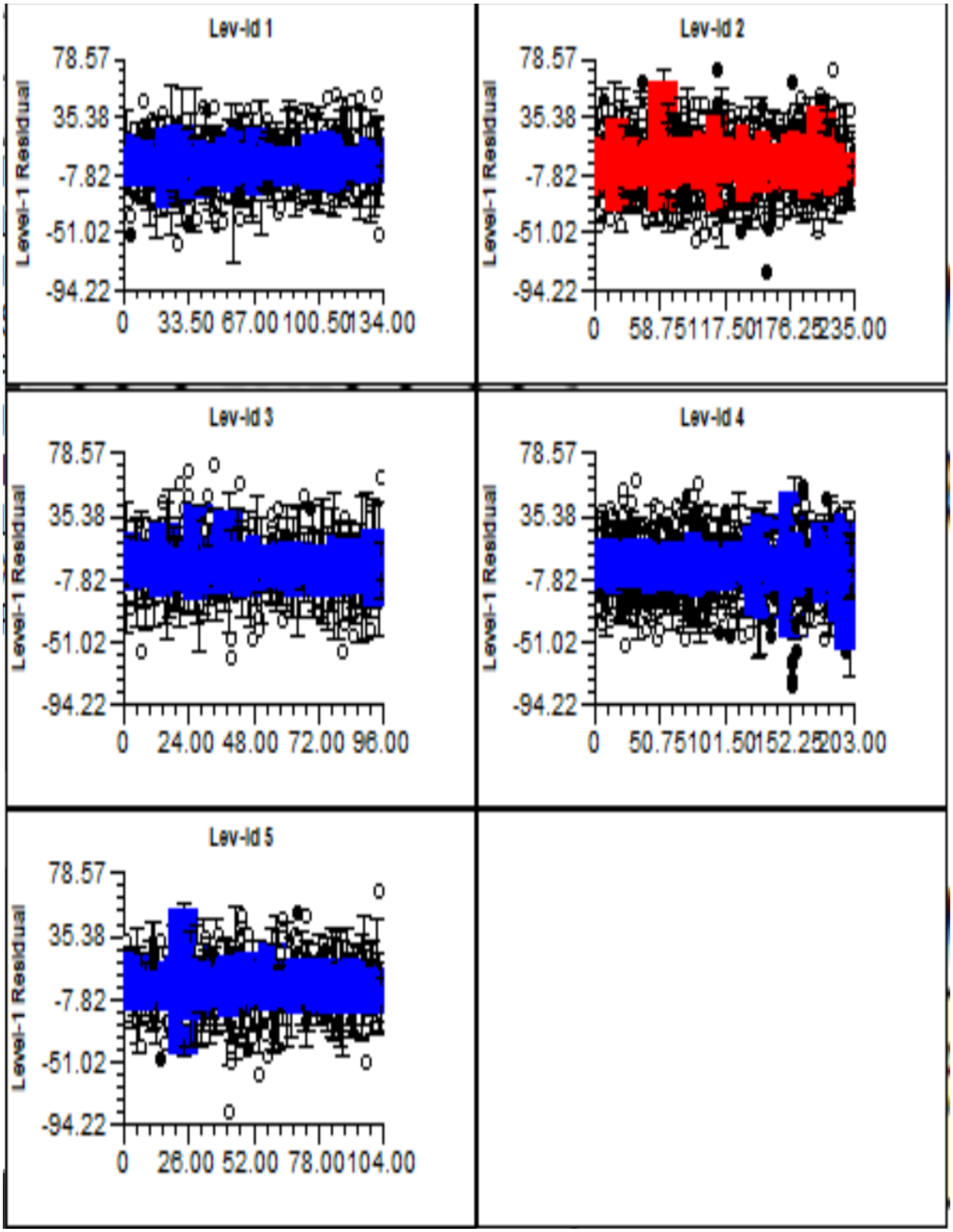


Figure A5.1c. Graph of Level 1 Residuals for 5 Semesters suggesting linearity among the included variables.

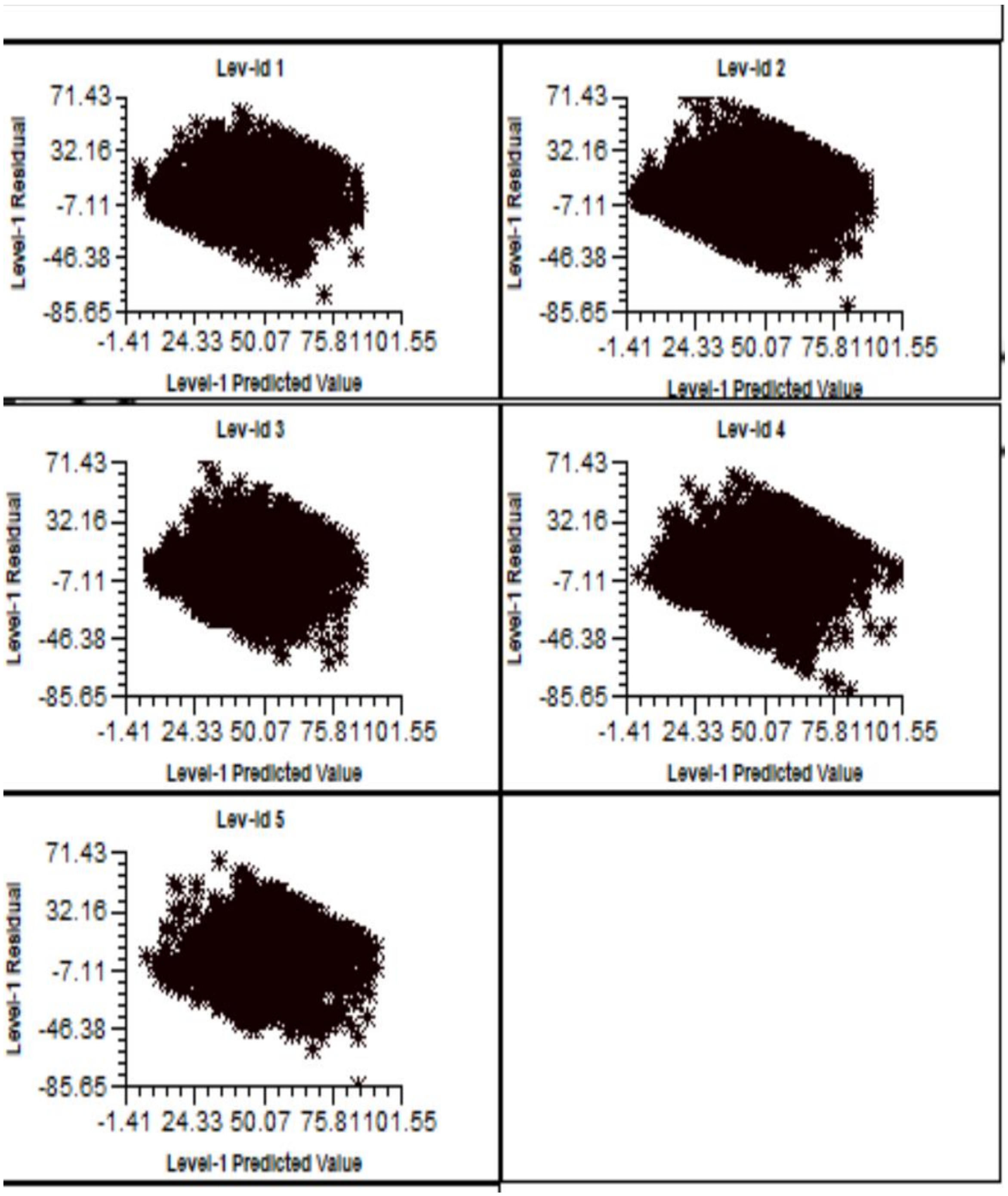


Figure A5.1d. Graph of Level 1 Residuals for 5 Semesters suggesting linearity among the included variables.

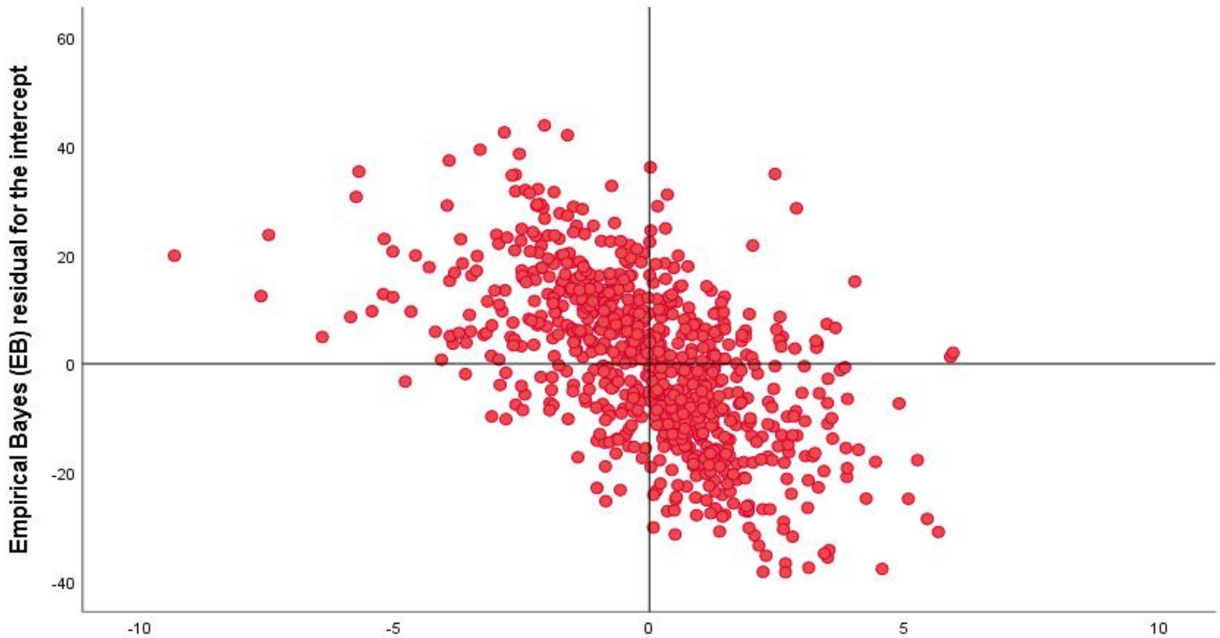


Figure A5.1e. Scatter diagram for residual vs. fitted value showing homogeneity of variance for Black.

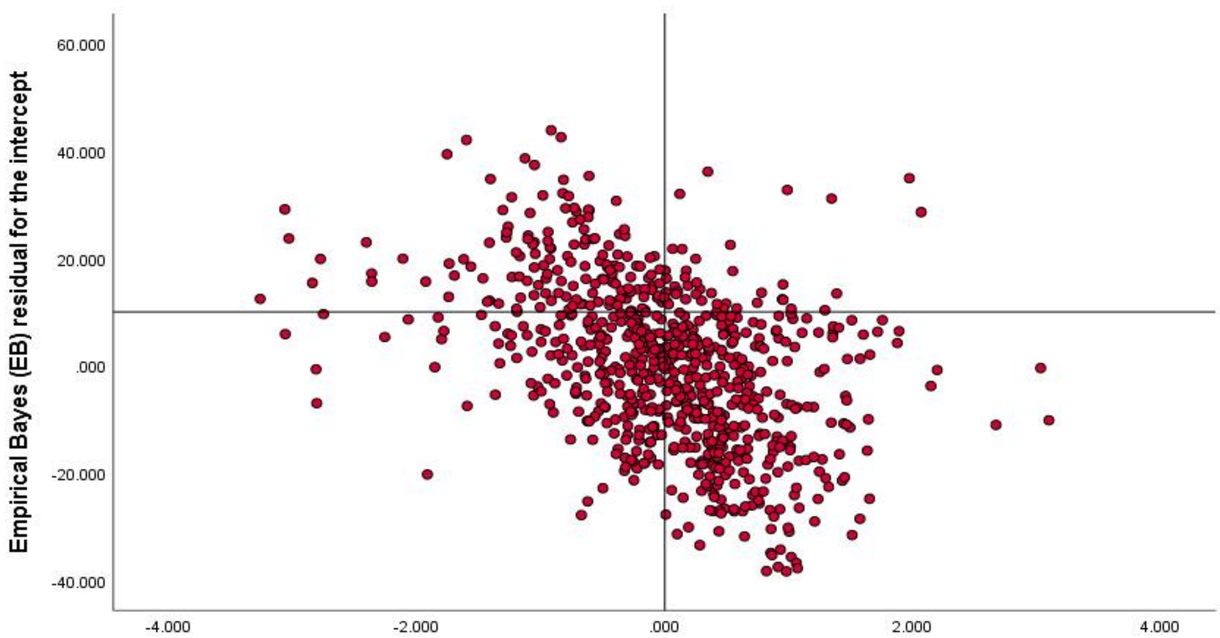


Figure A5.1f. Scatter diagram for residual vs. fitted value showing homogeneity of variance for Hispanics.

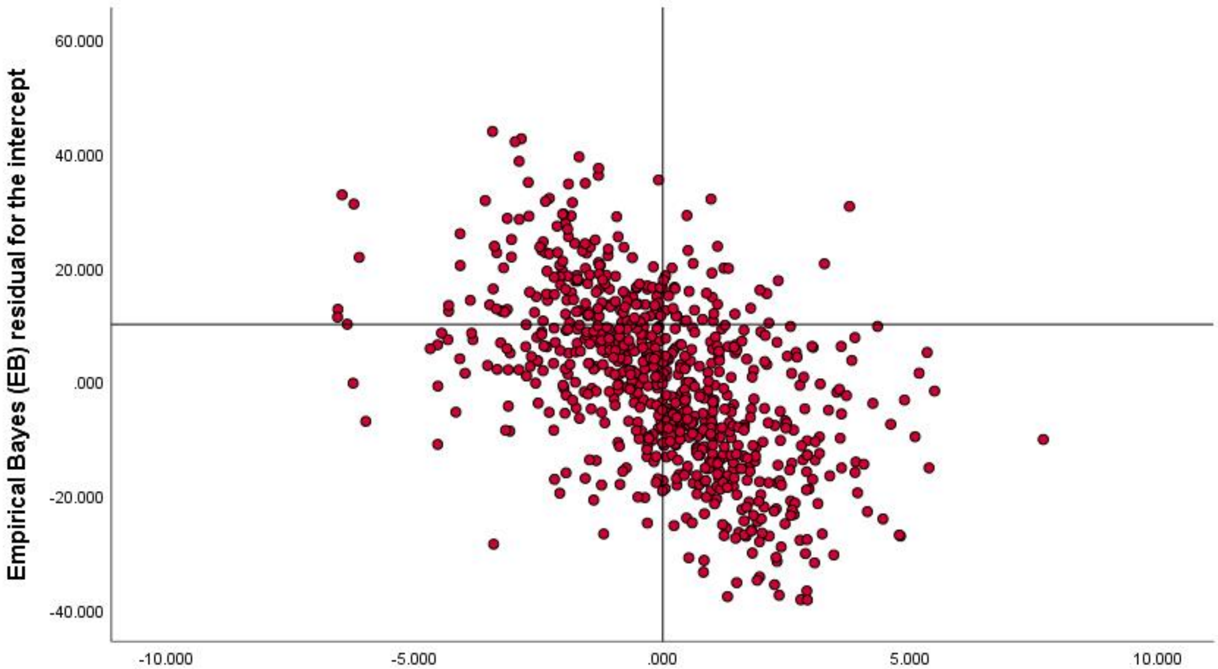


Figure A5.1g. Scatter diagram for residual vs. fitted value showing homogeneity of variance for FRPL.

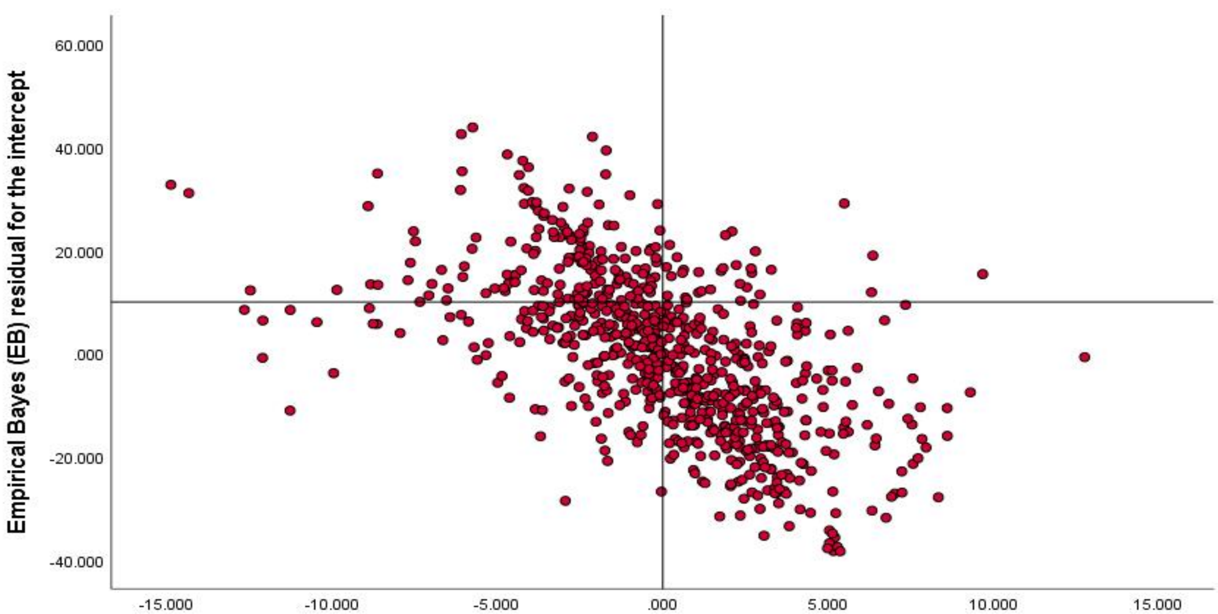


Figure A5.1h. Scatter diagram for residual vs. fitted value showing homogeneity of variance for EL.

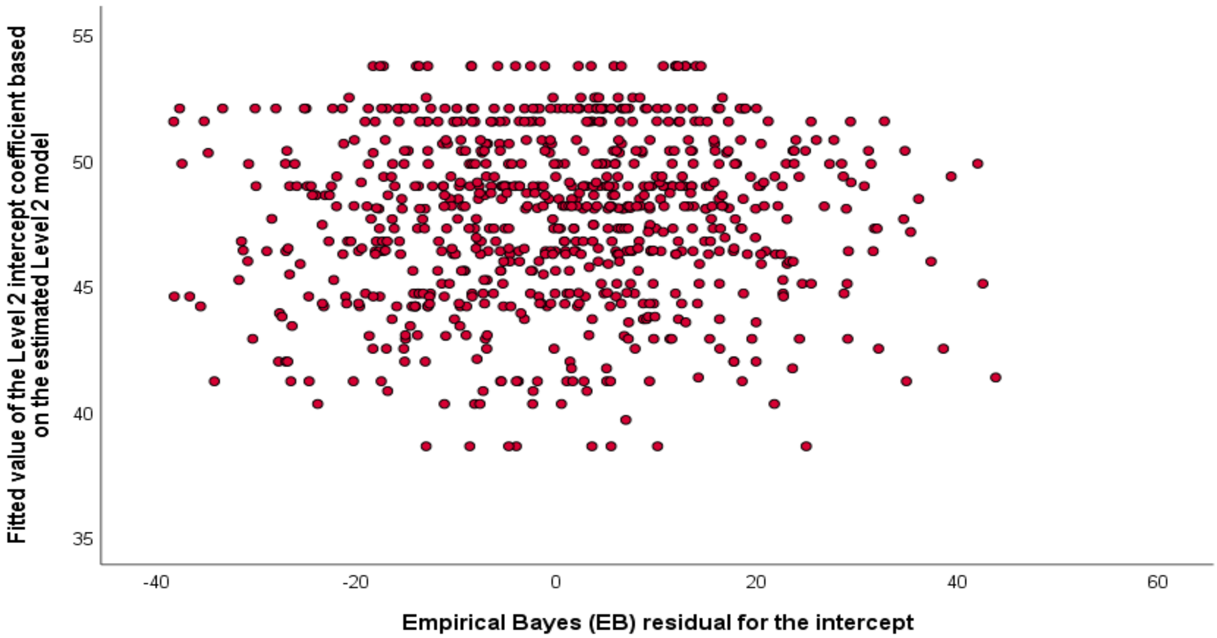


Figure A5.2a. Scatter diagram for intercept residual vs. fitted value showing Level 2 normality.

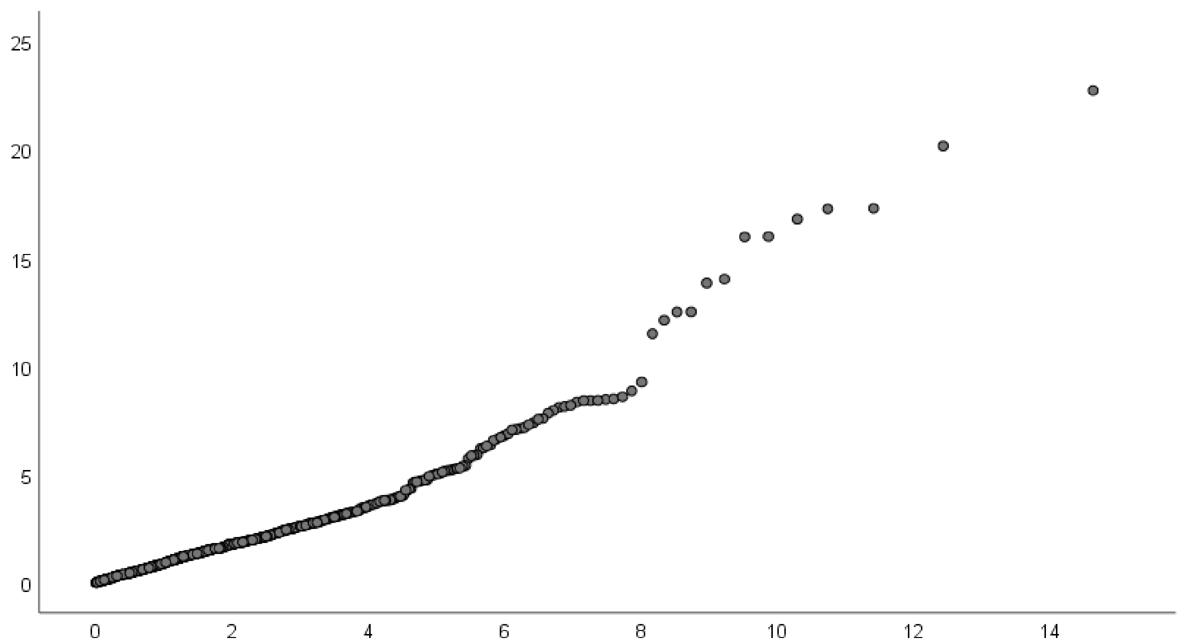


Figure A5.2b. Normal Q-Q plot showing Level 2 normality for each unit based on Order statistics, expected from a chi-square distribution for MIDST and Mahalanobis distance.

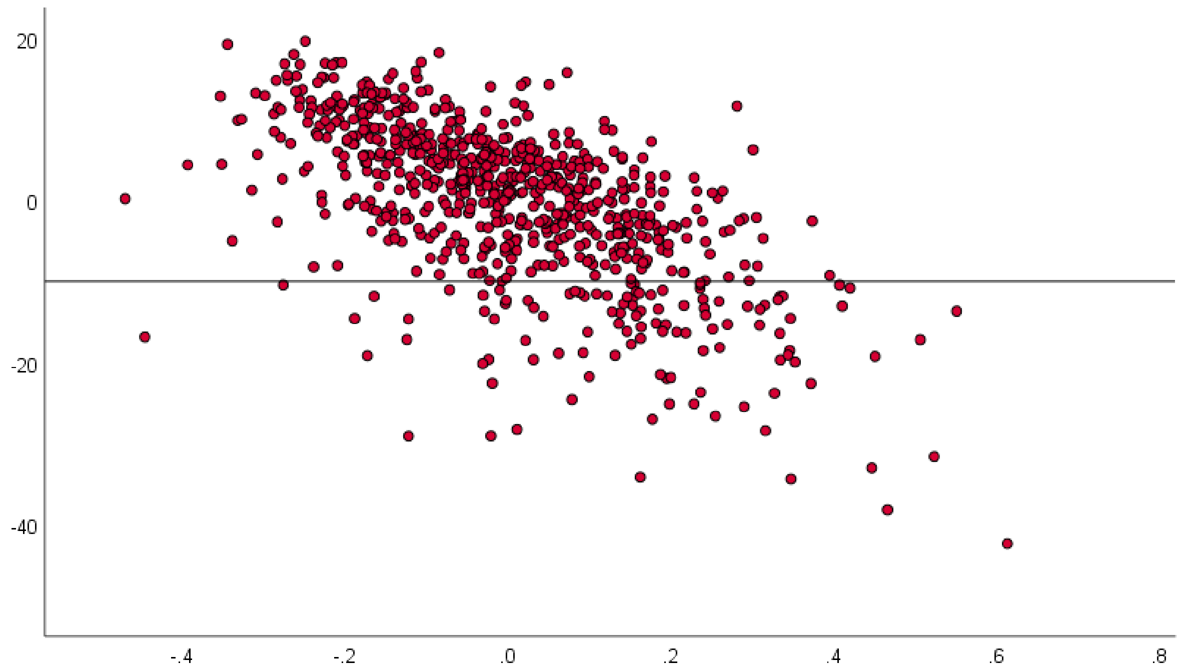


Figure A5.2c. Scatter diagram for residual versus fitted values showing homogeneity of variance for PRE_SCR.

LEVEL 3 RESIDUAL ANALYSIS

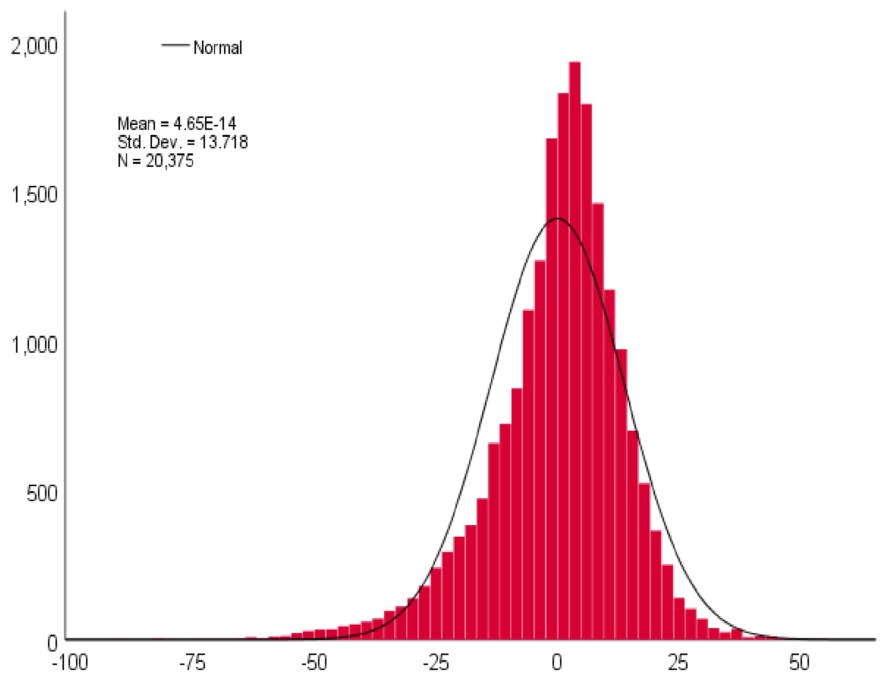


Figure A5.3a. Histogram with a normal curve for Empirical Bayes residual analysis for predicting semester model for PST level intercept.

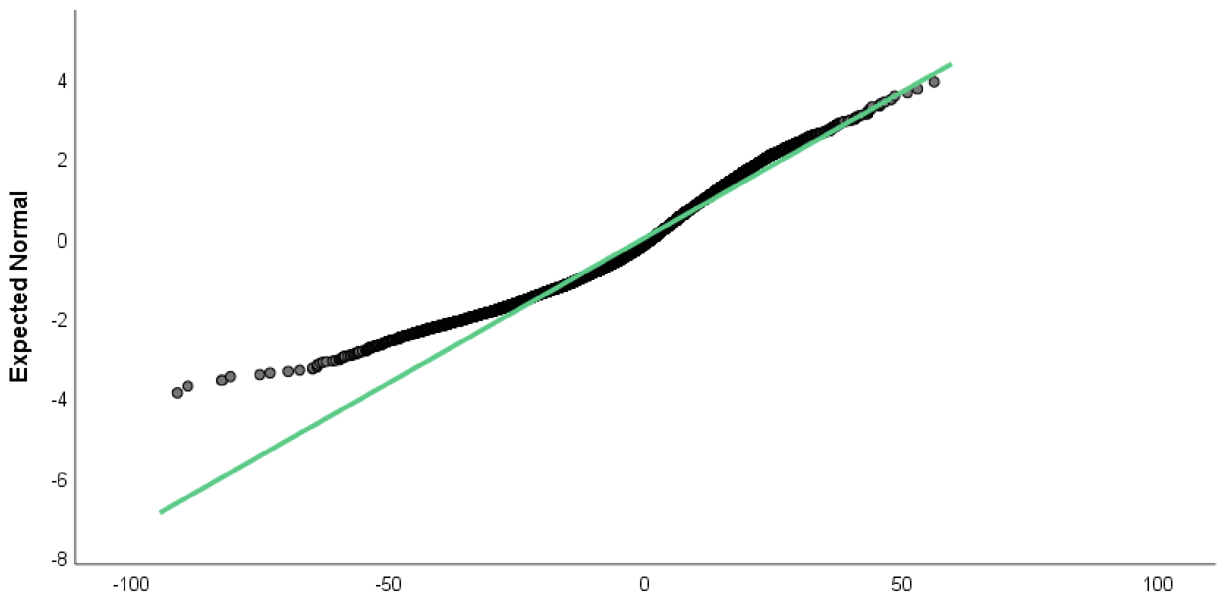


Figure A5.3b. Normal Q-Q plot of observed and fitted values showing Level-3 normality.

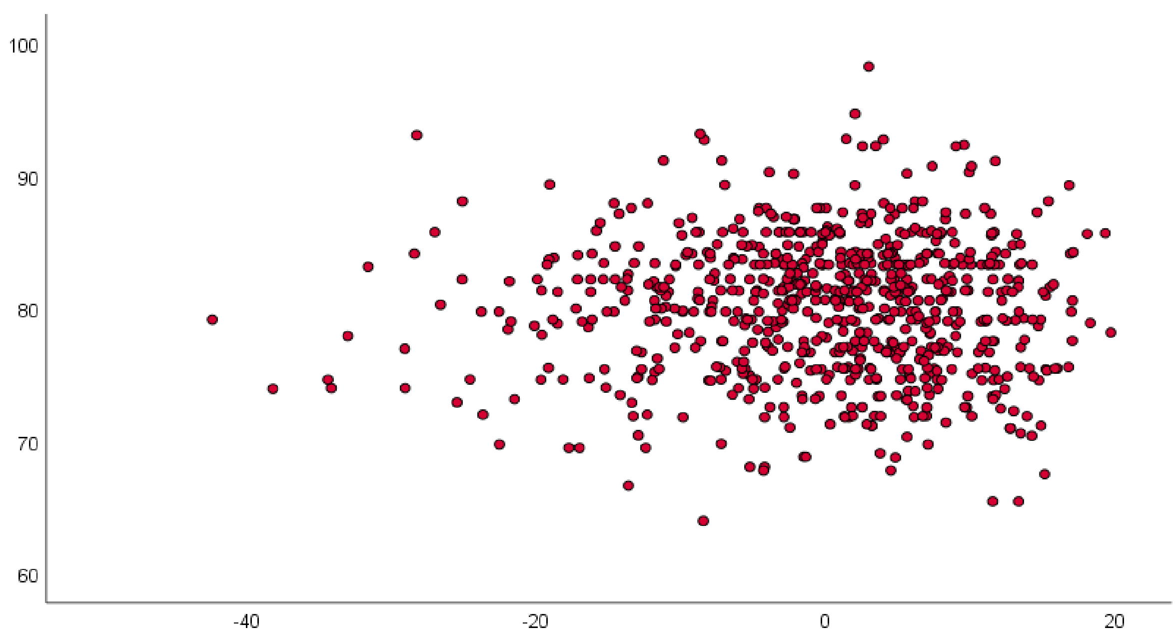


Figure A5.3c. Scatter diagram for residual versus fitted values showing homogeneity of variance.

APPENDIX F
GRAPHS OF TWO-WAY SIMPLE EFFECTS OF LEVEL 1 AND LEVEL 2 VARIABLES ON STUDENTS' ENGLISH LEARNER (EL) STATUS IN PREDICTING PRE- & POST TEST SCORES

STUDENTS' EL STATUS*PSTS' AREA OF STUDY (MAJOR)

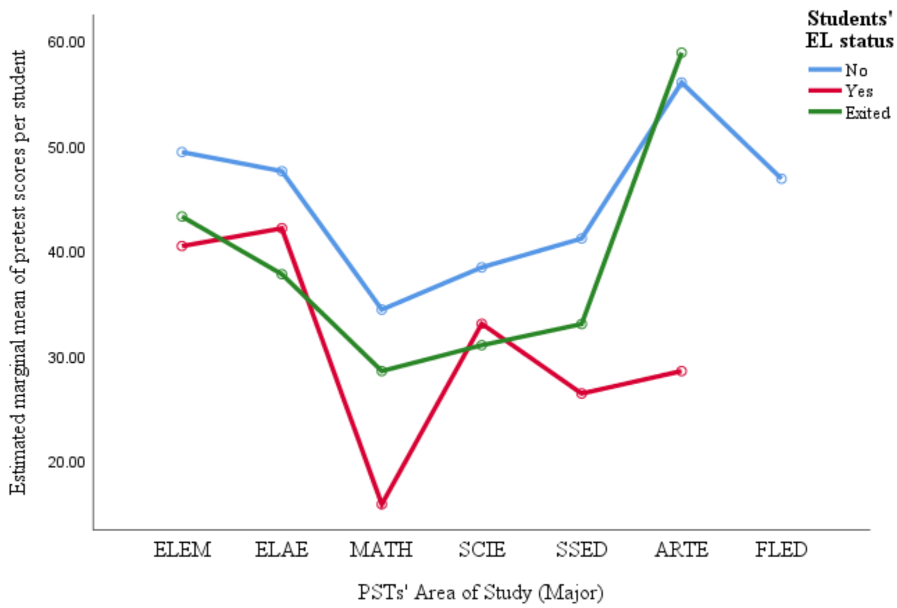


Figure A6.1a. The profile plot of pretest scores showing interactional effect of students' EL status and PSTs' area of study.

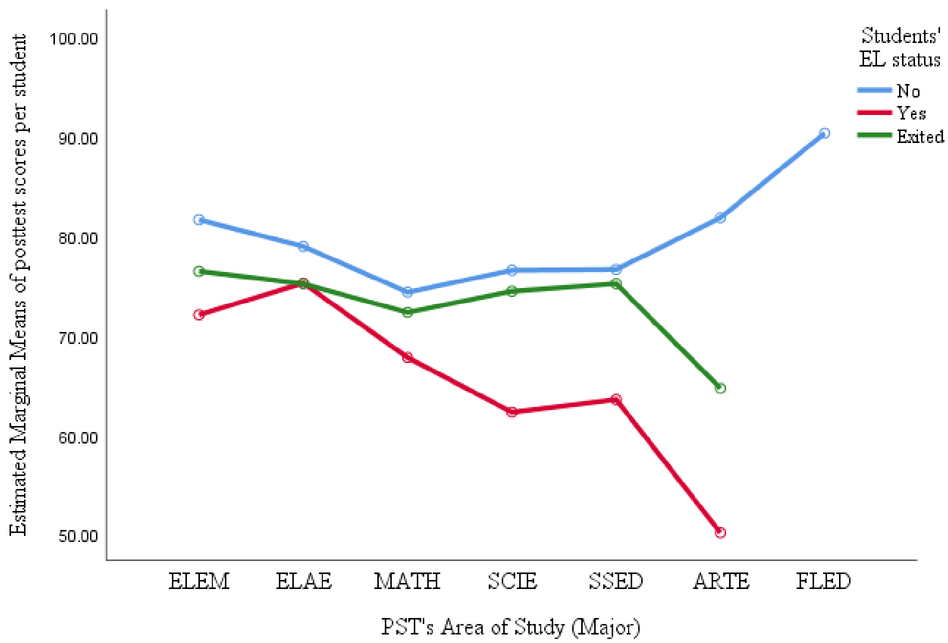


Figure A6.1b. The profile plot of posttest scores showing interactional effect of students' EL status and PSTs' area of study.

STUDENTS' EL STATUS*CLASS SIZE

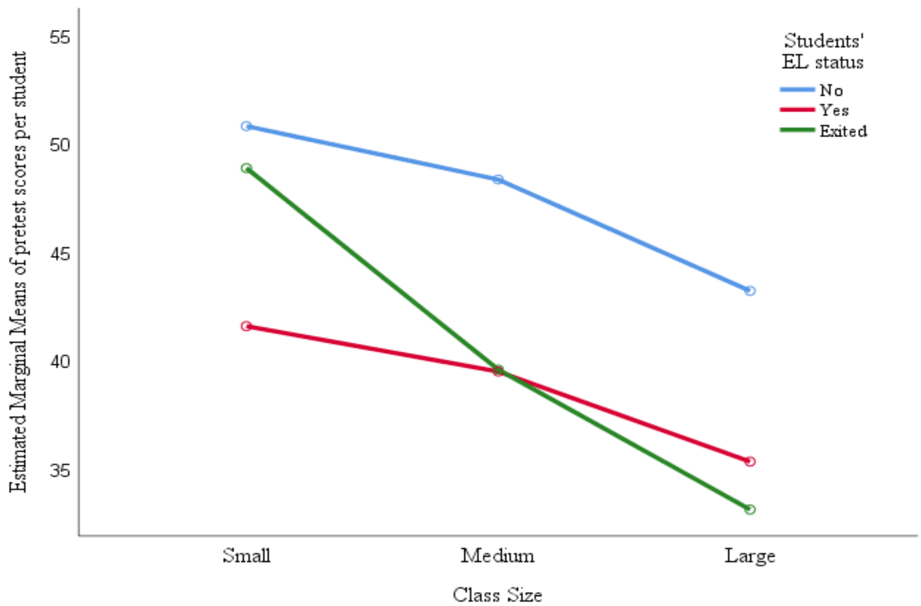


Figure A6.2a. The profile plot of pretest scores showing interactional effect of students' EL status and their class sizes.

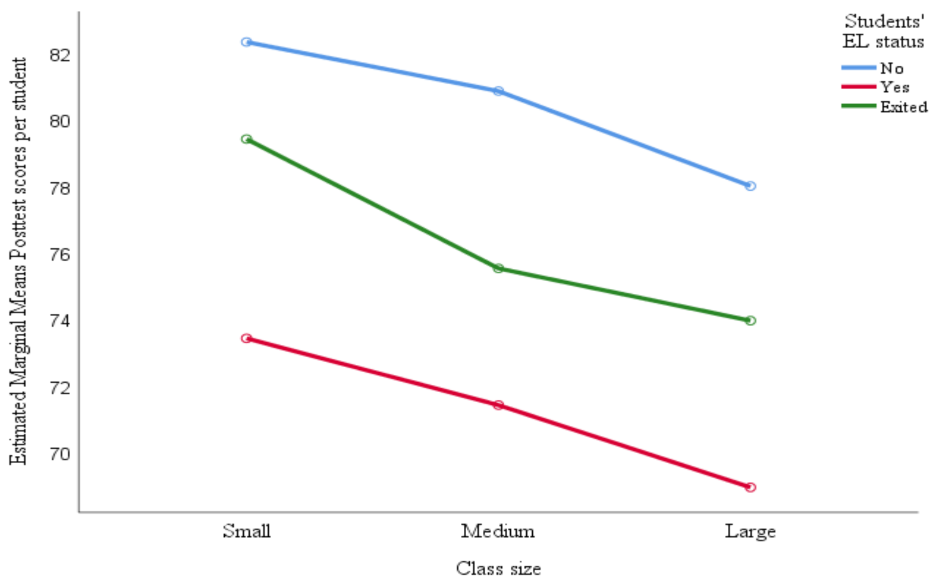


Figure A6.2b. The profile plot of posttest scores showing interactional effect of students' EL status and their class sizes.

STUDENTS' EL STATUS*CONTENT AREA (SUBJECT)

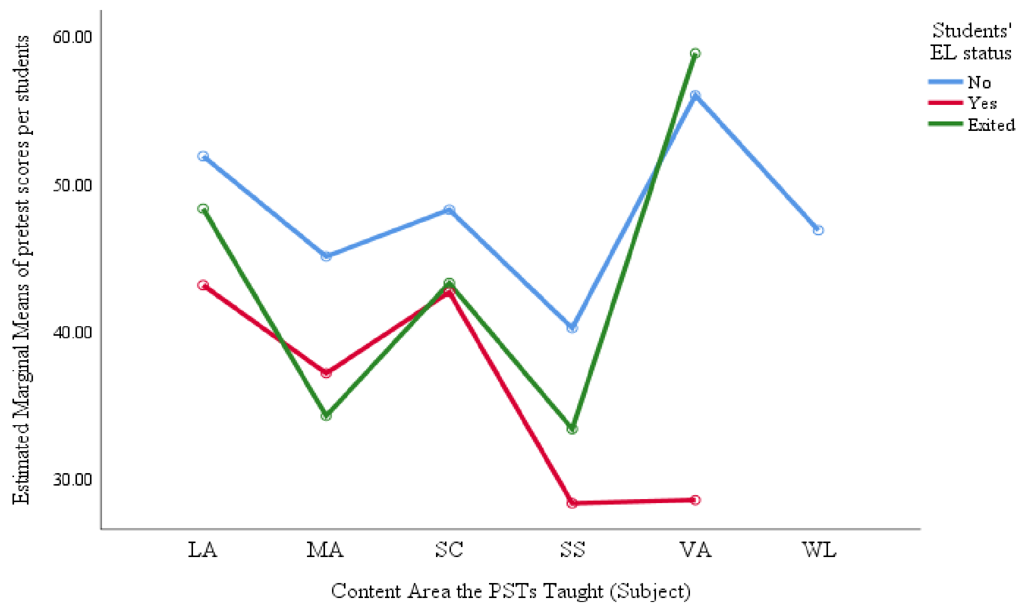


Figure A6.3a. The profile plot of pretest scores showing interactional effect of students' EL status and the content area.

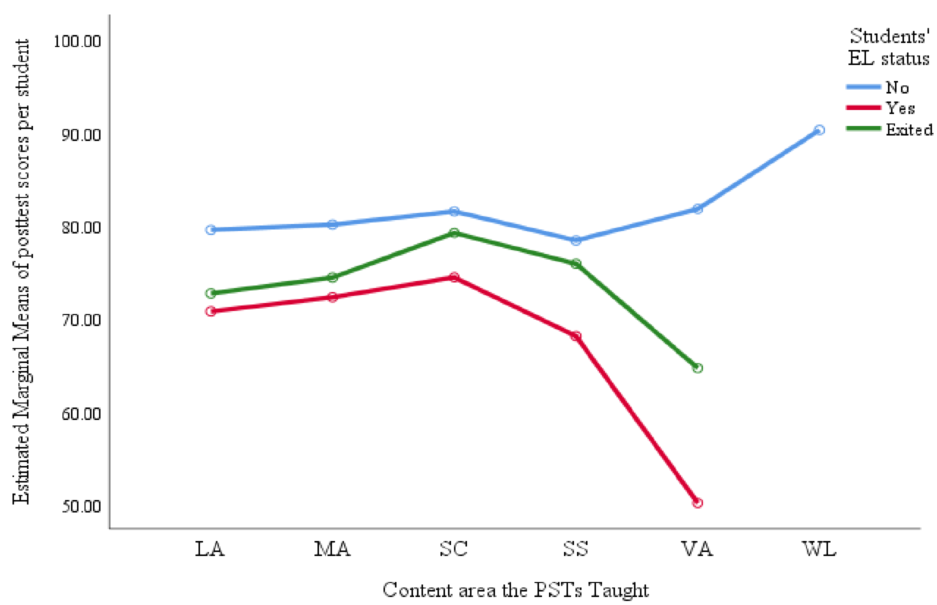


Figure A6.3b. The profile plot of posttest scores showing interactional effect of students' EL status and the content area.

STUDENTS' EL STATUS*GRADE LEVEL PST TAUGHT

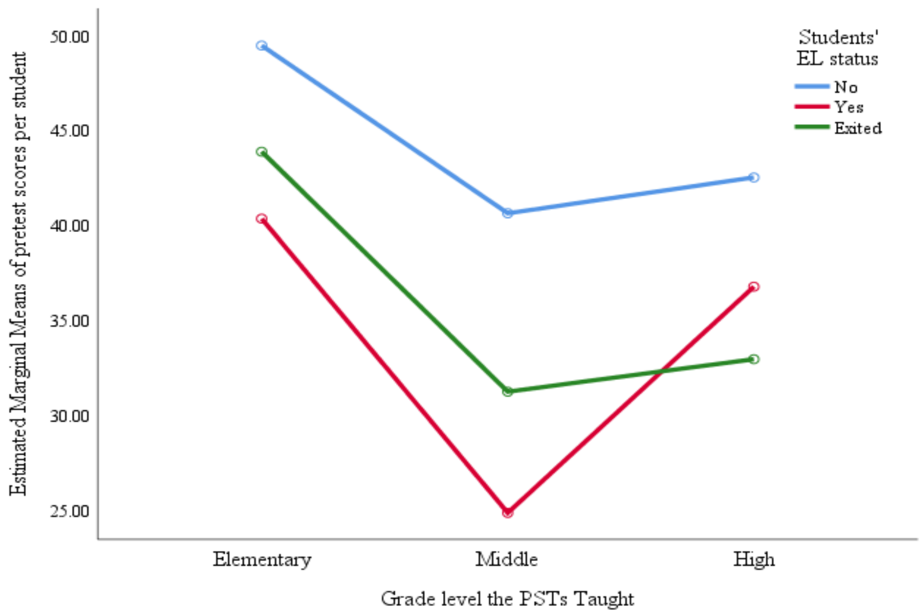


Figure A6.4a. The profile plot of pretest scores showing interactional effect of students' EL status and the grade level the PSTs taught.

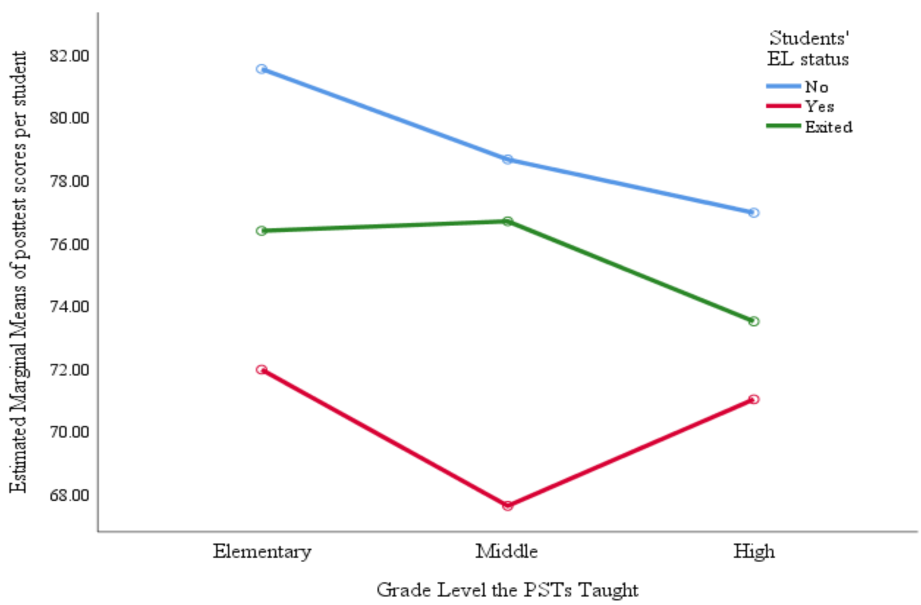


Figure A6.4b. The profile plot of posttest scores showing interactional effect of students' EL status and the grade level the PSTs taught.

STUDENTS' EL STATUS*STUDENTS' GENDER

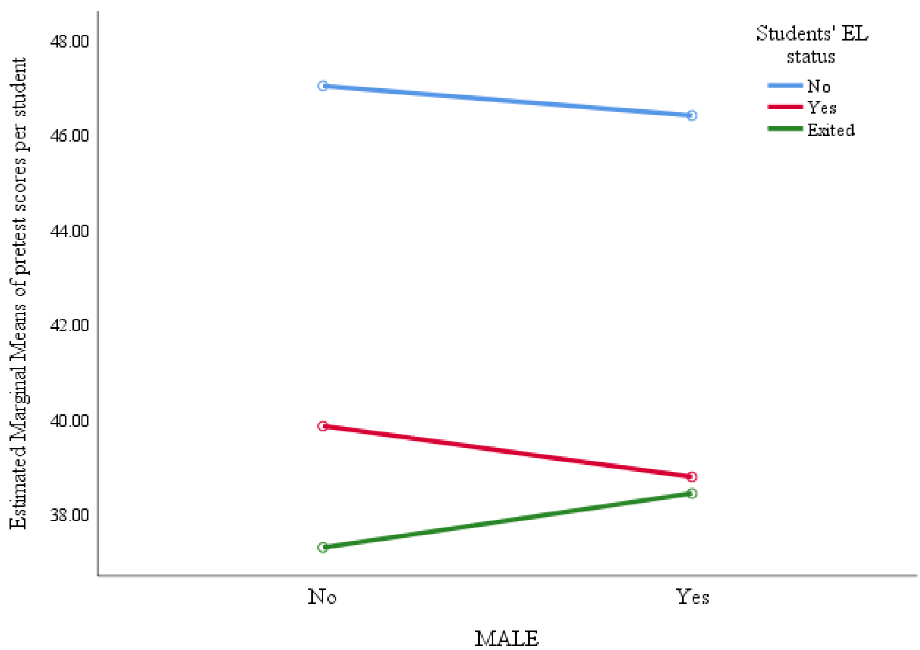


Figure A6.5a. The profile plot of pretest scores showing interactional effect of students' EL status and their gender.

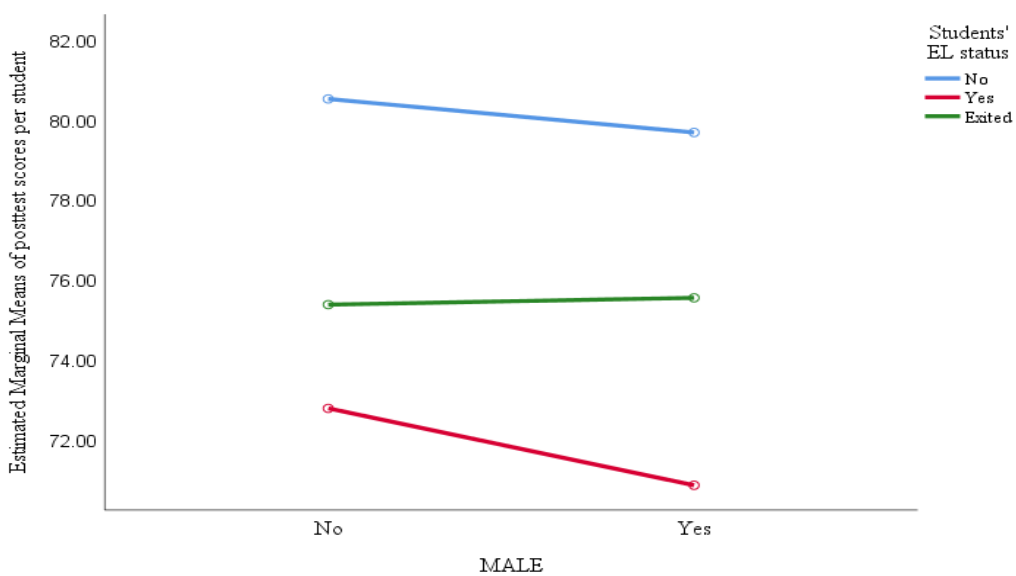


Figure A6.5b. The profile plot of posttest scores showing interactional effect of students' EL status and their gender.

STUDENTS’ EL STATUS*STUDENTS’ MINORITY STATUS

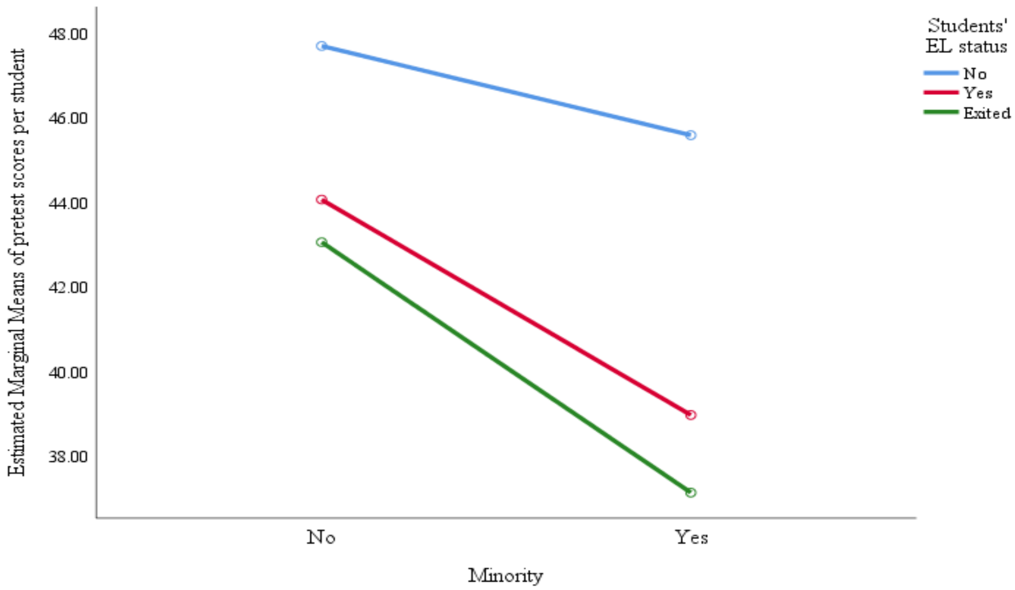


Figure A6.6a. The profile plot of pretest scores showing interactional effect of students’ EL status and their minority status.

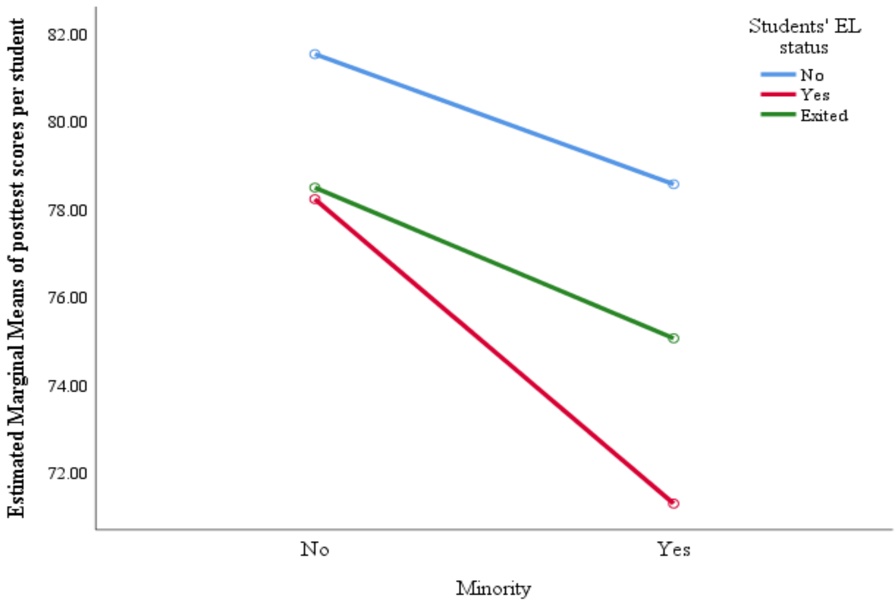


Figure A6.6b. The profile plot of posttest scores showing interactional effect of students’ EL status and their minority status.

STUDENTS' EL STATUS*STUDENTS' SOCIOECONOMIC STATUS

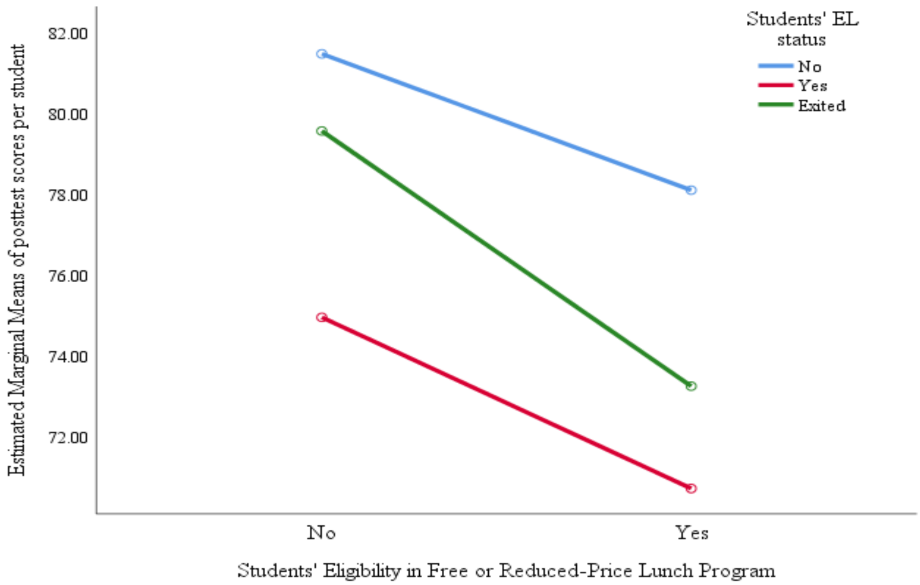


Figure A6.7a. The profile plot of posttest scores showing the interactional effect of students' EL status and their socioeconomic status.

LIST OF REFERENCES

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of labor Economics, 25*(1), 95-135.

Anderson, S., Medrich, E., & Fowler, D. (2007). Which achievement gap?. *Phi Delta Kappan, 88*(7), 547-550. https://doi.org/101177/003172170708800716

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment, 8*(3), 231-257.

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*(8), 448-452.

Allday, R. A., Neilsen-Gatti, S., & Hudson, T. M. (2013). Preparation for inclusion in teacher education pre-service curricula. *Teacher Education and Special Education, 36*(4), 298-311.

Akalin, S., & Sucuoglu, B. (2015). Effects of Classroom Management Intervention Based on Teacher Training and Performance Feedback on Outcomes of Teacher-Student Dyads in Inclusive Classrooms. *Educational Sciences: Theory and Practice*, *15*(3), 739-758.

Anderson, S., Medrich, E., & Fowler, D. (2007). Which achievement gap? *Phi Delta Kappan*, *88*(7), 547-550.

Andrews, S. J. (2003). Teacher language awareness and the professional knowledge base of the L2 teacher. *Language Awareness, 12*(2), 81-95.

Andrews, S. J. (1999). Why do L2 teachers need to ‘know about language’? Teacher metalinguistic awareness and input for learning. *Language and Education, 13*(2/3), 161-177.

Anthony, A., & Walshaw, M. (2007). Effective pedagogy in mathematics/pangarau. Best Evidence Synthesis Iteration [BES]. Wellington, New Zealand: Ministry of Education.

Arens, A. K., & Morin, A. J. (2016). Relations between teachers’ emotional exhaustion and students’ educational outcomes. *Journal of Educational Psychology*, *108*(6), 800-813.

Atkinson, D. (2011). A sociocognitive approach to second language acquisition: How mind, body, and world work together in learning additional languages. In *Alternative approaches to second language acquisition* (pp. 155-178). Routledge.

Athanases, S. Z., Bennett, L. H., & Wahleithner, J. M. (2013). Fostering data literacy through preservice teacher inquiry in English language arts. *The Teacher Educator*, *48*(1), 8-28.

Baca, L., & Escamilla, K. (2002). Educating teachers about language. In C. Adger, C. E. Snow, & D. Christian (Eds.), *What teachers need to know about language* (p. 71-83). Washington, DC, and mcHenry, IL: Center for Applied Linguistics and Delta System.

Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *International Journal of Epidemiology*, 1336-1339. doi: 10.1093/ije/dyu080

Ball, S. J. (2009). Privatizing education, privatizing education policy, privatizing educational research: Network governance and the 'competition state.' *Journal of Education Policy, 24*(1), 83-99.

Ballantyne, K. G., Sanderman, A. R., & Levy, J. (2008). *Educating English Language Learners: Building Teacher Capacity*. Washington, DC: National Clearinghouse for English Language Acquisition.

Baker, J. A., Kamphaus, R. W., Horne, A., & Winsor, A. P. (2006). Evidence for population-based perspectives on children's behavioral adjustment and needs for service delivery in schools. *School Psychology Review, 35*, 31–46.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Banks, J. A. (Ed.). (2012). *Encyclopedia of Diversity in Education (Vols. 1-4).* Thousand Oaks, CA: Sage.

Banks, J. A., & Banks, C. A. M. (Ed.). (2004). *Handbook of research on multicultural education (2nd ed.).* San Francisco, CA: Jossey-Bass.

Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies, 2*(2), 285-324.

Barwell, R. (2005). What works? Research into practice. A research-into practice series produced by a partnership between the Literacy and Numeracy Secretariat and the Ontario Association of Deans of Education. *Research Monogram*, *34*, 1-4.

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.

Benton, J. E., Powell, D., DeLine, M. A., Sautter, A., Talbut, M. H., Bratberg, W., & Cwick, S. (2012). The teacher work sample: A professional culminating activity that integrates general studies objectives. *The Journal of General Education, 61*(4), 369-387.

Berkowitz, R., Moore, H., Astor, A. R., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research, 87*(2), 425-469.

Boero, P., Douek, N., & Ferrari, P. L. (2008). 12 Developing mastery of natural language: Approaches to some theoretical aspects of mathematics. *Handbook of international research in mathematics education, 262.*

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational researcher*, *34*(6), 3-15.

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.

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.

Braun, H. I. (2005). Using student progress to evaluate teachers: A Primer on value-added models. *Educational Testing Services*. Retrieved from https://eric.ed.gov/?id=ED529977

Brincks, A. M., Enders, C. K., Llabre, M. M., Bulotsky-Shearer, R. J., & Prado, G. (2017). Centering Predictor Variables in Three-Level Contextual Models. *Multivariate Behavioral Research*, *52*(2), 149-163.

Bristor, V. J., Pelaez, G. M., & Crawley, S. (2000). An integrated elementary education/ESOL teacher preparation program. *Action in Teacher Program, 22*(2), 25-32. doi: 10.1080/01626620.2000.10463002

Broman, G. D., Grigg, J., & Hanselman, P. (2016). An effort to close achievement gaps at scale through self-affirmation. *Educational Evaluation and Policy Analysis, 38*(1), 21-42.

Brown, C. (2017). *Quality education in Idaho: A case study of academic achievement in three high-poverty rural schools* Available from ProQuest Dissertations & Theses Full Text: The Humanities and Social Sciences Collection. Retrieved from https://search.proquest.com/docview/1936354246

Bunch, G. C. (2013). Pedagogical language knowledge: Preparing mainstream teachers for English learners in the new standards era. *Review of Research in Education*, *37*(1), 298-341.

Burn, K., Mutton, T., Thompson, I., Ingram, J., McNicholl, J., & Firth, R. (2016). The impact of adopting a research orientation towards use of the Pupil Premium Grant in preparing beginning teachers in England to understand and work effectively with young people living in poverty. *Journal of Education for Teaching*, *42*(4), 434-450.

Byrnes, D. A., Kiger, G., & Manning, M. L. (1997). Teachers' attitudes about language diversity. *Teaching and teacher education, 13*(6), 637-644.

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.

Chen, X., & Yeping, L. (2008). Language proficiency and mathematics learning. *School Science and Mathematics, 108*(3), 90-94.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Discussion of the American statistical association's statement (2014) on using value-added models for educational assessment. *Statistics and Public Policy, 1*(1), 111-113.

Coady, M., & Escamilla, K. (2005). Audible voices, visible tongues: Exploring social realities in Spanish-speaking students' writing. *Language Arts, (82)*, 462-471.

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.

Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 375-395). Dordrecht, the Netherlands: Kluwer.

Cochran-Smith, M., Stringer Keefe, E., Chang, W. C., & Cummings Carney, M. (2018). NEPC Review: 2018 State Teacher Policy Best Practices Guide (National Council on Teacher Quality, March 2018).

Cochran-Smith, M., & Villegas, A. M. (2015). Framing teacher preparation research: An overview of the field, part I. *Journal of Teacher Education, 66*(1), 7-20.

Cochran-Smith, M. (2001). Constructing outcomes in teacher education. *Education Policy Analysis Archives*, *9*(11), 1-56.

Cohen, J. (1992). Statistical power analysis. *Current Direction in Psychological Sciences, 1*(3), 98-101.

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.

Cooper, J. A., & Schleser, R. (2006). Closing the achievement gap: Examining the role of cognitive developmental level in academic achievement. *Early Childhood Education Journal, 33*(5), 301-306.

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.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A Reassessment1. *Applied linguistics*, *2*(2), 132-149.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire* (Vol. 23). Multilingual Matters.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher, 44*(2), 132-137. doi:10.1177/0022487105285962

Darling-Hammond, L. (2014a). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator,* 4-44.

Darling-Hammond, L. (2014b). Want to close the achievement gap? Close the teaching gap. *American Educator, 38*(4), 14.

Darling-Hammond, L. (2008). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57*(2), 120-138. doi: 10.117/0022487105283796

Darling-Hammond, L. (2006a). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57*(20), 120-138. doi: 10.1177/0022487105283796

Darling-Hammond, L. (2006b). Constructing 21st-century teacher education. *Journal of Teacher Education, 57*(3), 300-314.

Darling-Hammond, L. (2004). What happens to a dream deferred; research on families, schools, and communities; engaging life a funds of knowledge approach. In J. A. Banks, C A M Banks (Ed.), *Handbook of research on multicultural education* (2nd ed., pp. 607-630). San Francisco, CA: Jossey-Bass.

Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation. *Journal of Teacher Education, 53*(4), 286-302.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *ERIC Clearinghouse on Assessment and Evaluation, 8*(1), 1-44.

de Boer, A., Pijl, S. J., & Minnaert, A. (2011). Regular primary schoolteachers' attitudes towards inclusive education: A review of the literature. *International Journal of Inclusive Education, 15*(3), 331-353.

de Jong, E. J., Harper, C. A., & Coady, M. R. (2013). Preparing mainstream teachers for CLD students: Enhancing the knowledge and skills that teachers of CLDs must have. *Theory into Practice*, 52(2), 89.97.

de Oliveira, L. C., & Athanases, S. Z. (2007). Graduates' reports of advocating for English language learners. *Journal of Teacher Education, 58*(3), 202-215.

Decker, P. T., Rice, J. K., & Moore, M. T. (1997). *Education and the economy: An indicators report*. National Center for Education Statistics, US Department of Education, Office of Educational Research and Improvement. https://nces.ed.gov/pubs97/97269.pdf

Denner, P. R., Salzman, S. A., & Bangert, A. W. (2001). Linking teacher assessment to student performance: A benchmarking, generalizability, and validity study of the use of teacher work samples. *Journal of Personnel Evaluation in Education*, *15*(4), 287-307.

Denner, P. R., Salzman, S. A., & Newsome, J. D. (2001). Selecting the qualified: A standards-based teacher education admission process. *Journal of Personnel Evaluation in Education*, *15*(3), 165-180.

Dong, Y. R. (2004). Preparing secondary subject area: Teachers to teach linguistically and culturally diverse students. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 77*(5), 202-206.

Doung, M. T., Badaly, D., Liu, F. F., Schwartz, D., & McCarty, C. A. (2016). Generational differences in academic achievement among immigrant youths: A meta-analytic review. *Review of Educational Research, 86*(1), 3-41.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. Russell Sage Foundation.

Ellis, R. (Eds.). (2015). *The study of second language acquisition* (2nd ed.). Oxford, NY: Oxford University Press.

Echevarria, J., MaryEllen. V., & Short, D. (2008). *Making content comprehensible for English learners: The SIOP model* (4th ed.). New York, NY: Pearson Education Inc.

Education Commission of the States. (2014). How is an "English language learner" defined in state policy? Retrieved October 1, 2019, from http://ecs.force.com/mbdata/mbquestNB2?rep=ELL1402.

Enders, C. K., Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.

Faltis, C., Arias, M. B., & Ramirez-Marin, F. (2010). Identifying relevant competencies for secondary teachers of English learners. *Bilingual Research Journal, 33*(3), 307-328.

Fatih, K. (2017). Understanding the roles of SES, ethnicity, and discipline infractions in students' standardized test scores. *European Journal of Alternative Education Studies*, *2*(1), 114-127.

Feiman-Nemser, S. (2018). What does research tell us about educating mainstream teachers to work with ELLs? *The Educational Forum, 82*(2), 227.

Feiman-Nemser, S. (2008). Teacher learning: How do teachers learn to teach. *Handbook of research on teacher education: Enduring questions in changing contexts*, *3*, 697-705.

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103*(6), 1013-1055.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education, 38*(4), 460-507.

Fillmore, L. W., & Snow, C. E. (2000). What teachers need to know about language. U. S. Department of Education, Office of Education Research and Improvement. Retrieved from: https://files.eric.ed.gov/fulltext/ED444379.pdf

Forlin, C. (2010). Developing and implementing quality inclusive education in Hong Kong: Implications for teacher education. *Journal of Research in Special Educational Needs, 10*, 177-184.

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, 12*(1), 117-136.

Francis, D., & Stephens, A. (Eds.). (2018). *English learners in STEM subjects: Transforming classrooms, schools, and lives.* Washington, DC: The National Academies Press.

Fredman, T. (2004). Teacher work sample methodology: Implementation and practical application in teacher preparation. *Action in Teacher Education, 26*(1), 3-11.

Freeman, D., & Johnson, K. E. (2005). Toward linking teacher knowledge and student learning. *Second language teacher education: International perspectives*, 73-95.

Freeman, D., & Johnson, K. E. (1998). Reconceptualizing the knowledgebase of language teacher education. *TESOL Quarterly, 32*(3), 397-417.

Furgione, B., Evans, K., Ghimire, N., Thripp, R., & Russell, W. B. III. (2018). What's on your plate? Correlating subsidized lunch and proficiency on the civics end-of-course assessment. *Educational Practice and Theory, 40*(1), 51-66.

Furlong, John, Marilyn Cochran-Smith, and Marie Brennan, eds. (2013). *Policy and politics in teacher education: International perspectives*. Routledge.

Galguera, T. (2011). Participant structures as professional learning tasks and the development of pedagogical language knowledge among preservice teachers. *Teacher Education Quarterly, 38*(1), 85–106.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction (8th ed.).* Boston, MA: Allyn & Bacon.

Gandara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs. *Policy Analysis for California Education, PACE (NJ1)*.

García, E., Arias, M. B., Harris Murri, N. J., & Serna, C. (2010). Developing responsive teachers: A challenge for a demographic reality. *Journal of Teacher Education, 61*(1-2), 132-142.

Gass. S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *International Association of Applied Linguistics, 19*(1), 3-17. https://doi.org/10.1075/aila.19.03gas

Gass. S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). New York, NY: Taylor & Francis.

Genesee, F., Lindholm-Leary, K., Saunders, W. M., & Christian, D. (2009). *Educating English language learners: Synthesis of research evidence*. New York, NY: Cambridge University Press.

Gehrke, R. S., & Cocchiarella, M. (2013). Preservice special and general educators' knowledge of inclusion. *Teacher Education and Special Education, 36*(3), 204-216.

Ghimire, N., Purmensky, K. L., & Regmi, S. (2020). Does infused ESOL teacher preparation program hold promise towards narrowing the English learner achievement gap? A multilevel analysis of the One-Plus model [Manuscript submitted for publication]. College of Arts and Humanities, University of Central Florida

Ghimire, N., & Regmi, S. (in press). Do school and school district influence students' performance on civics end-of-course assessment? A multilevel analysis. *Florida Journal of Educational Research.*

Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children, 85*(3), 329-346.

Girod, G. R. (Ed.). (2002). *Connecting teaching and learning: A handbook for teacher educators on teacher work sample methodology*. Washington, DC: AACTE Publications.

Goddard, R. D., Sweetland, S. R. & Hoy, W. K. (2000). Academic emphasis on urban elementary schools and student achievement in reading and mathematics: A multilevel analysis. *Educational Administration Quarterly, 36*(5), 683-702.

Goldenberg, C. (2008). Teaching English language learners: What the research does-and does not-say. *Open Educational Resources-ESED 5234*. Retrieved from https://digitalcommons.georgiasouthern.edu/esed5234-master/27

Gollnick, D. (2002). Incorporating linguistic knowledge in standards for teacher performance. In C. Adger, C. E. Snow, & D. Christian (Eds.), *What teachers need to know about language* (p. 103-112). Washington, DC, and McHenry, IL: Center for Applied Linguistics and Delta Systems.

Gonzalez, J. E., & Darling-Hammond, L. (2000). Programs that prepare teachers to work effectively with students learning English. *ERIC Clearinghouse on languages and Linguistics*.

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.

Gorman, S. (2010). *National Assessment of Educational Progress: An overview of NAEP*. Washington, D. C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Dept. Of Education. Retrieved from https://nces.gov/pubsearch/pubsinfo.asp?pubid=2010468

Govoni, J. (2011). The evolution, experiences, and lessons learned with ESOL infusion in Florida. *The Tapestry Journal, 3*(2), 1-6.

Graham, K. J., & Fennel, F. (2001). Principles and standards for school mathematics and teacher education: Preparing and empowering teachers. *School Science and Mathematics, 101*(6), 319-327.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39*(1), 59-68.

Grissmer, D., Flanagan, A., Treisman, P. U., Fuller, E. J., & Meyer, R. H. (2001). Searching for indirect evidence for the effects of statewide reforms. *Brookings Papers on Education Policy, 4*, 181-229.

Gross, L. A., Fitts, S., Goodson-Espy, T., & Clark, A. M. (2010). Self as teacher: Preliminary role identification of the potential teaching candidate. *Australian Journal of Teacher Education, 35*(2), 1-19.

Guo, L., Peng, J., & Xie, Q. (2018). *Maximum likelihood estimation-based regression for multivariate calibration, 189,* 316-321.

Guo, S., & Hussey, D. (1999). Analyzing longitudinal rating data: A three-level hierarchical linear model. *Social Work Research, 23*(4), 258-269.

Hadwin, A. F., & Winne, P. H. (2012). Study strategies have meager support: A review with recommendations for implementation. *The Journal of Higher Education, 67*(6), 692-715.

Hahs-Vaughn, D. L. (2016). *Applied multivariate statistical concept.* New York, NY: Routledge.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2^{nd} ed.). Washington DC: SAGE Publication.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed, a silver bullet. *Journal of Marketing Theory and Practice, 19*, 139-151.

Hammerness, K., Darling-Hammond, L., & Bransford, J., with Berliner, D., Cochran-Smith, M., McDonald, M., & Zeichner, K. (2005). How teachers learn and develop. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (p. 358-389). San Francisco: Jossey-Bass.

Hanushek, E. A., & Rivkin, S. G. (2006). School quality and the Black-White achievement gap [working paper]. *National Bureau of Economic Research.* Retrieved from: https://uscupstate.libguides.com/c.php?g=452037&p=3086943

Harvey, M. W., Yssel, N., Bauserman, A. D., Merbler, J. B. (2010). Preservice teacher preparation for inclusion: An exploration of higher education teacher-training institutions. *Remedial and Special Education, 31,* 24-33.

Hennin, J. E., DeBruin-Parecki, A., Hawbaker, B. W., Nielsen, C. P., Joram, E., & Gabrele, A. J. (2005). The teacher work sample: A tool for scaffolding and assessing preservice teachers' early field experiences. *The Teacher Educator, 40*(3), 188-207. doi:10.1080/08878730509555360

Herman, M. R. (2009). The black-white-other achievement gap: Testing theories of academic performance among multiracial and monoracial adolescents. *Sociology of Education, 82*(1), 20-46.

Ho, A. D., & Kane, T. J. (2013). The reliability of classroom observations by school personnel. Research Paper. MET Project. *Bill & Melinda Gates Foundation.*

Hoffert, S. B. (2009). Mathematics: The universal language? *Mathematics Teacher, 103*(2), 130-139.

Hoffman, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management 23*(6), 723-744.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organization. *Journal of Management, 24*(5), 623-641.

Howey, K. R., & Zimpher, N. L. (1989). *Profiles of preservice teacher education: Inquiry into the nature of programs.* Sunny Press.

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, *58*(8), 702-714.

Hung, M., Smith, W. A., Voss, M. W., Franklin, J. D., Gu, Y., & Bounsanga, J. (2019). Exploring achievement gaps in school districts across the United States. *Education and Urban Society, 00*(0), 1-19.

Hunter, R. C., & Bartee, R. (2003). The achievement gap: Issues of competition, class, and race. *Education and Urban Society, 35*(2), 151-160.

Hutchinson, M. (2013). Bridging the gap: Preservice teachers and their knowledge of working with English language learners. *TESOL Journal*, *4*(1), 25–54.

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.

Israel, G. D. (1992). *Determining sample size*. Retrieved from: https://www.gjimt.ac.in/wp-content/uploads/2017/10/2_Glenn-D.-Israel_Determining-Sample-Size.pdf

Jeynes, W. H. (2015). A meta-analysis on the factors that best reduce the achievement gap. *Education and Urban Society, 47*(5), 523-554.

Jimenez-Silva, M., Olson, K., & Jimenez Hernandez, N. (2012). The confidence to teach English language learners: Exploring coursework's role in developing preservice teachers' efficacy. *The Teacher Educator, 47*(1), 9-28.

Johnson, T, & Wells, L. (2017). English language learner teacher effectiveness and the common core. *Education Policy Analysis Archives, 25*(23).

Jolly, G. M. (1969). Sampling methods for aerial censuses of wildlife populations. *East African Agricultural and Forestry Journal, 34*(1), 46-49.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives, 16*(4), 91-114.

Karabenick, S. A., & Noda, P. A. C. (2004). Professional development implications of teachers' beliefs and attitudes toward English language learners. *Bilingual Research Journal, 28*(1), 55-75.

Kazima, M., & Adler, J. (2006). Mathematical knowledge for teaching: Adding to the description through a study of probability in practice. *Pythagoras, 63*, 46-59.

Kindler, A. M. (2003). Visual culture, visual brain, and (art) education. *Studies in Art Education*, *44*(3), 290-296.

Klees, S. 2008. "A Quarter Century of Neoliberal Thinking in Education: Misleading Analyses and Failed Policies." *Globalisation, Societies, and Education, 6* (4), 311–348.

Klingner, J. K., & Artiles, A. J. (2003). When should bilingual students be in special education? *Educational Leadership, 61*(2), 66-71.

Kim, J. (2011). Relationships among and between ELL status, demographic characteristics, enrollment history, and school persistence. CRESST Report 810. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Koedel, C., & Betts, J. R. (2007). *Re-examining the role of teacher quality in the educational production function*. National Center on Performance Incentives, Vanderbilt, Peabody College.

Koh, J. H. L., Chai, C. S., & Lim, W. Y. (2017). Teacher professional development for TPACK-21CL: Effects on teacher ICT integration and student outcomes. *Journal of Educational Computing Research, 55*(2), 172-196.

Kontopantelis, E., White, I. R., Sperrin, M., & Buchan, L. (2017). Outcome-sensitive multiple imputation: a simulation study. *BMC Medical Research Methodology, 17*(2), 1-13. doi 10.1186/s12874-016-0281-5

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*(3), 643-680.

Korthagen, F. A., & Kessels, J. P. (1999). Linking theory and practice: Changing the pedagogy of teacher education. *Educational researcher*, *28*(4), 4-17.

Kotok, S. (2017). Unfulfilled potential: High-achieving minority students and the high school achievement gap in math. *High School Journal, 100*(3), 183-202.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation and Policy Analysis, 25*(3), 287-298.

Laarhoven, T. R. V., 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.

Lachapelle, C. P., Oh, Y., Shams, M. F., Hertel, J. D., & Cunningham, C. M. (2015). HLM modeling of pre/post-assessment results from a large-scale efficacy study of elementary engineering. In *122nd ASEE Annual Conference & Exposition: Making Value for Society*, American Society for Engineering Education, Seattle. Retrieved from https://pennstate.pure.elsevier.com/en/publications/hlm-modeling-of-prepost-assessment-results-from-a-large-scale-eff

Lavery, M. R., Nutta, J., & Youngblood, A. (2019). Analyzing student learning gains to evaluate differentiated teacher preparation for fostering English learners' achievement in linguistically diverse classrooms. *Journal of Teacher Education, 00(0),* 1-16.

Lavery, M. R. (2012). TWS GraphMaker [Computer software]. Orlando, FL: M. R. Lavery.

Lavy, I., & Mashiach-Eizenberg, M. (2009). The interplay between spoken language and informal definitions of statistical concepts. *Journal of Statics Education, 17(1).*

Lee, O., Maerten-Rivera, J., Penfield, R. D., LeRoy, K., Secada, W. G. (2008). Science achievement of English language learners in urban elementary schools: Results of a first-year professional development intervention. *Journal of Research in Science Teaching, 45*(1), 31-52.

Lewis-Beck, M. S., Bryman, A., & Liao, T. F (Eds.). (2004). *The SAGE encyclopedia of social science research methods*. Sage Publication. Assessed from:

http://methods.sagepub.com/reference/the-sage-encyclopedia-of-social-science-research-methods/n619.xml

Lightbown, P. M., & Spada, N. (2006). *How languages are learned* (3rd ed.). Oxford: Oxford University Press.

Linquanti, R., & Cook, H. G. (2013). Toward a "Common Definition of English Learner": A Brief Defining Policy and Technical Issues and Opportunities for State Assessment Consortia. *Council of Chief State School Officers*.

Linton, T. H., & Kester, D. (2003). Exploring the achievement gap between white and minority students in Texas. *Education Policy Analysis Archives, 11*(10), 1-19.

Linville, H. A. (2015). ESOL teachers as advocates: An important role? *TESOL Journal, 7*(1), 98-131.

Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school. *Educational Researcher, 46*(2), 103-109.

Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis*, *36*(4), 457-475.

Lohman, D. F., & Gambrell, J. L. (2012). Using nonverbal tests to help identify academically talented children. *Journal of Psychoeducational Assessment*, *30*(1), 25-44.

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*, *52*(4), 275-296.

Lucas, T., & Villegas, A. M. (2013). Preparing linguistic responsive teachers: Laying the foundation in preservice teacher education. *Theory into Practice*, 52(20), 98-109.

Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education. *Journal of Teacher Education, 59*(4), 361-373.

Markos, A. M. (2012). Mandated to learn, guided to reflect: Pre-service teachers’ evolving understanding of English language learners. *Issues in Teacher Education*, *21*(1), 39–57.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample size for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, (1*(3), 86-92.

Maulana, R., Helms-Lorenz, M, & van de Grift, W. (2015). A longitudinal study of induction on the acceleration of growth in teaching quality of beginning teachers through the eyes of their students. *Teaching and Teacher Education, 51*, 225-245.

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

McCoach, D. B. (2010). Hierarchical linear modeling. *The reviewer's guide to quantitative methods in the social sciences*, 123-140.

McGrath, K. F., & Van Bergen, P. (2015). Who, when, why and to what end? Students at risk of negative student–teacher relationships and their outcomes. *Educational Research Review*, *14*, 1-17.

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.

Mead, J. F., & Eckes, S. E. (2018). *How school privatization opens the door for discrimination.* National Education Policy Center. https://files.eric.ed.gov/fulltext/ED591982.pdf

Meaney, T., & Flett, K. (2006). Learning to read in mathematics classrooms. *Australian Mathematics Teacher*, *62*(2), 10-16.

Medina, A. L., Hathaway, J. I., & Pilonieta, P. (2015, Spring). How preservice teachers' study abroad experiences lead to changes in their perceptions of English language learners. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *25*, 73–91.

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

Moallem, M. (1998). An expert teacher's thinking and teaching and instructional design models and principles: An ethnographic study. *Educational technology research and development*, *46*(2), 37-64.

Mun, R. U., Langley, S. D., Ware, S., Gubbins, E. J., Siegle, D., Callahan, C. M., McCoach, D. B., & Hamilton, R. (2016). *Effective practices for identifying and serving English learners in gifted education: A systematic review of the literature.* National Center for Research on Gifted Education. https://eric.ed.gov/?id=ED602387

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90-100.

National Center for Education Statistics (NCES). (2015). *Achievement gaps.* https://nces.ed.gov/nationsreportcard/studies/gaps/

National Council for Accreditation of Teacher Education. (2008). *Professional standards for the accreditation of teacher preparation institutions*. Massachusetts, NW: National Council for Accreditation of Teacher Education.

National Council of Teachers of English, The James R. Squire Office for Policy Research. (2008). *English Language Learners: A policy research brief.* Retrieved from http://www.ncte.org/library/NCTEFiles/Resources/PolicyResearch/ELLResearchBrief.pdf

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics*.

Niday, D. (2011). Infusing EL content and instruction into English education courses. *The Tapestry Journal, 3*(2), 19-25.

Noland, A., & Richards, K. (2015). Servant teaching: An exploration of teacher servant leadership on student outcomes. *Journal of the Scholarship of Teaching and Learning*, *15*(6), 16-38.

Nutta, J. W., Strebel, C., Mihai, F. M., Creveocoeur-Bryant, E., & Mokhtari, K. (2018). *Show, tell, build: Twenty key instructional tools and techniques for educating English Learners.* Cambridge, MA: 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.). Cambridge, MA: Harvard Education Press.

Nutta, J. W., Mokhtari, K., & Strebel, C. (2012). *Preparing every teacher to reach English learners: A practical guide for teacher educators*. Cambridge, MA: Harvard Education Press.

O'Conner, R., Abedi, J., & Tung, S. (2012). *A descriptive analysis of enrollment and achievement among English language learner students in Delaware.* (Issues and Answers Report, REL 2012-No. 132). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from http://ies.ed.gov/ncee/edlabs.

Opdenakker, M., Damme, J. V., De Fraine, D. F., Landeghem, G. V., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement, 13*(4), 399-427.

Peugh, J. L., & Heck, R. H. (2017). Conducting three-level longitudinal analyses. *Journal of Early Adolescence, 37*(1), 7-58.

Portes, R. P. (2005). *Dismantling educational inequality. A cultural approach to closing the achievement gap*. New York, NY: Peter Lang.

Posselt, J. R., Jaquette, O., Bielby, R., & Bastedo, M. N. (2012). Access without equity: Longitudinal analyses of institutional stratification by race and ethnicity, 1972-2004. *American Educational Research Journal, 49*(6), 1074-1111.

Powers, P. J. (1992, October). *The effect of special education coursework upon the preparation of pre-service teachers* (Eric Document Reproduction Service No. ED 377 183). Paper presented at the annual meeting of the Northern Rocky Mountain Educational Research Associations, Custer, SD.

Pu, C. (2012). Narrative inquiry: Preservice teachers' understanding of teaching English learners. *AILACTE Journal*, *9*(1), 1–18.

Raubaugh, L. (2019). Cultural responsiveness and K-12 English learners: Exploring policy and implementation fidelity [Doctoral dissertation, University of Central Florida]. https://stars.library.ucf.edu/etd/6563/

Repko, A. F., & Szostak, R. (2017). *Interdisciplinary research: Process and theory*. SAGE Publications, Inc.

Rheams, T. A., & Bain, S. K. (2005). Social interaction interventions in an inclusive era: Attitudes of teachers in early childhood self-contained and inclusive settings. *Psychology in the Schools*, *42*(1), 53-63.

Richardson, V. (Ed.). (2002). *Handbook of research on teaching*. American Educational Research Association.

Sanders, S. M., Durbin, J. M., Anderson, B. G., Fogarty, L. M., Giraldo-Garcia, R. J., Voight, A. (2018). Does a rising school climate lift all boats? Differential associations of perceived climate and achievement for students with disabilities and limited English proficiency. *School Psychology International, 39*(6), 646-662. https://doi.org/10.1177/0143034318810319

Santiago, R., Martinez, V., & Peter, A. (2016). *ELL procedural handbook: Multilingual student education services*. Orange County Public Schools. https://www.ocps.net/UserFiles/Servers/Server_54619/File/Departments/Multilingual/FINAL%20%20ELLProceduralHandbook-revised%20September%209%202016.pdf

Sato, M. (2014). What is the underlying conception of teaching of the edTPA? *Journal of Teacher Education, 65*(5), 421-434.

Schalock, H. D., & Myton, D. V. (1988). A new paradigm for teacher licensure: Oregon's demand for evidence of success in fostering learning. *Journal of Teacher Education, 39*(6), 8-16.

Schmidt, P. C. (2005). From teacher candidates to ESL ambassadors in teacher education. *TESL-EJ: Teaching English as a Second or Foreign Language, 9*(2), 11.

Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teachers College Record, 112*(5), 1492-1528.

Schuetz, P., & Schuetz, P. (2005). UCLA community college review: Campus environment- A missing link in studies of community college attrition. *Community College Review*, *32*(4), 60-80.

Selwyn, D. (2007). Highly quantified teachers: NCLB and teacher education. *Journal of Teacher Education*, *58*(2); 124-137.

Settlage, J., Gort, M., & Ceglie, R. J. (2014). Mediated Language Immersion and Teacher Ideologies: Investigating Trauma Pedagogy within a" Physics in Spanish" Course Activity. *Teacher Education Quarterly*, *41*(3), 47-66.

Sheppard, B. (2006). Class size in K-12 schools [Review of research evidence prepared for the Teacher Allocation Commission Newfoundland and Labrador]. https://www.researchgate.net/profile/Bruce_Sheppard/publication/269105937_Class_Size_in_K-12_Schools_A_Review_of_the_Research_Evidence/links/548100d40cf22525dcb6 06fc.pdf

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4-14.

Siegel, M. A. (2014). Developing preservice teachers’ expertise in equitable assessment for English learners. *Journal of Science Teacher Education*, *25*(3), 289–308.

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.

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Snijders, T. A., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Stobaugh, R. R., Tassell, J. L., & Norman, A. D. (2012). Improving preservice teacher preparation through the teacher work sample: Exploring assessment and analysis of student learning. *Action in Teacher Education, 32*(1), 39-53.

Student [Def. 1]. In *Cambridge Dictionary Online*, Retrieved October 1, 2019, from https://dictionary.cambridge.org/us/dictionary/english/student

Subedi, B. R. (2005). *A demonstration of the Three-Level hierarchical generalized linear model applied to educational research* (Identifier No. FSU_migr_etd-1521) [Doctoral Dissertation, Florida State University]. FSU's Digital Repository.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In Gass, S. and Madden, C., editors, Input in second language acquisition. Rowley, MA: Newbury House, 235-252.

Sykes, G., & Dibner, K. (2009). Fifty Years of Federal Teacher Policy: An Appraisal. *Center on Education Policy*.

Tarone, E., & Allwright, D. (2005). Second language teacher learning and student second language learning: Shaping the knowledge base. In D. J. Tedick (Eds.), *Second language teacher education: International perspectives*, (p. 5-23). New York, NY: Routledge.

TESOL International Association. (2019). Standards for Initial TESOL Pre-K-12 teacher preparation programs. *Alexandria, VA: Author*.

Thomas, L., & Krebs, C. J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America, 78*(2), 128-139.

Thurlow, M. L., Albus, D. A., & Lazarus, S. S. (2017). *2014-2015 publicly reported assessment results for students with disabilities and ELs with disabilities.* National Center on Educational Outcome. https://eric.ed.gov/?id=ED581570

Townsend, T., & Bates, R. (eds.). (2007). *Handbook of teacher education: Globalization, standards, and professionalism in times of change.* Dordrecht, the Netherlands: Springer.

Trappes-Lomax, H. R., & Ferguson, G. (Eds.). (2002). *Language in language teacher education* (Vol. 4). John Benjamins Publishing.

Troff, B. (2014). Folk belief theory, the rigor fap, and the achievement gap. *The educational forum, 78*(2), 174-189.

United States Department of Agriculture, Food and Nutrition Service. (2017). *The national school lunch program.* Retrieved from https://fns-prod.azureedge.net/sites/default/files/resource-files/NSLPFactSheet.pdf

The University of Southern Queensland. (2016). English Language Proficiency. Retrieved October 1, 2019, from https://policy.usq.edu.au/documents/161460PL#2.

Uro, G., & Barrio, A. (2013). English language learners in America's great city schools: Demographics, achievement and staffing. Council of the Great City Schools.

Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal, 89*(3), 410-426.

Valdés, G., Bunch, G., Snow, C., & Lee, C. (2005). Enhancing the development of students' language(s). In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for changing world: What teachers should learn and be able to do* (p. 126-168). San Francisco, CA: Jossey Bass.

Valdés, G., & Geoffrion-Vinci, M. (2011). 28 Heritage Language Students: The Case of Spanish. *The handbook of Hispanic sociolinguistics, 598*.

Van Lier, L. (2004). The semiotics and ecology of language learning. *Utbildning & Demokrati*, *13*(3), 79-103.

VanPatten, B., & Williams, J. (Eds.). (2007). *Theories in second language acquisition: An introduction*. Mahwah, NJ: Erlbaum.

VanPatten, B. & Benati, A. G. (2015). *Key terms in second language acquisition*. New York, NY: Bloomsbury Publishing Plc.

Villegas, A. M., de La Mora, K. S., Martin, A. D., & Mills, T. (2018). Preparing future mainstream teachers to teach English language learners: A review of empirical literature. *The Educational Forum, 82*(2), 138-155.

Virtue, D. C. (2009). Using "ESOL rounds" to prepare middle level candidates for work with English language learners. *Middle Grades Research Journal*, *4*(1), 1–24.

Wade, S. E., Fauske, J. R., & Thompson, A. (2008). Prospective teachers' problem solving in online peer-led dialogues. *American Educational Research Journal*, *45*(2), 398–442.

Walker, A., Shafer, J., & Iams, M. (2004). Not in my classroom: Teacher attitudes towards English language learners in the mainstream classroom. *NAEB Journal of Research and Practice, 21*(1), 130-160.

Walqui, A. (2008). Quality teaching for English Learners: The development of adolescent English Language Learners' potential through a pedagogy of high challenge and high support. WestEd. Retrieved January 12, 2008.

Walton, P., Baca, L., & Escamilla, K. (2002). A national study of teacher education: Preparation for diverse student populations. *Berkeley, CA: Center for Research on Education, Diversity, and Excellence (CREDE).*

Weber, E. (2009). Quantifying student learning: How to analyze assessment data. *Ecology 101,* 501-511.

Welch-Ross, M. K. (2010). *Language diversity, school learning, and closing achievement gaps* [Workshop summary]. Washington, D. C.: The National Academies Press.

Welsh, R. O. (2018). Student mobility, segregation, and achievement gaps: Evidence from Clark County, Nevada. *Urban Education, 53*(1), 55-85.

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

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). Teacher Preparation Research. *Current Knowledge, Gaps, and Recommendations*. Washington: Center for the Study of Teaching and Policy.

Wixom, M. A. (2015). ECS and National Experts Examine: State-Level English Language Learner Policies. *Education Commission of the States*.

Woltman, H., Feldstain, A., Mckay, C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology, 8*(1), 52-69.

Chen, X., & Li, Y. (2008). Language proficiency and mathematics learning. *School Science and Mathematics*, *108*(3), 90-94.

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.

Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, *86*(4), 981-1015.

Zeichner, K. (2014). The struggle for the soul of teaching and teacher education in the USA. *Journal of Education of Teaching*, *40*(5), 551-568.

Zinth, J. D. (2013). Career/technical education: Not your father's vocational education. *The progress of education reform*, *14*(1), 1-7.