

Bowling Green State University

ScholarWorks@BGSU

School of Educational Foundations, Leadership
and Policy Faculty Publications

College of Education and Human Development

1-12-2018

Analyzing Student Learning Gains to Evaluate Differentiated Teacher Preparation for Fostering English Learners' Achievement in Linguistically Diverse Classrooms

Matthew Ryan Lavery

Bowling Green State University, mlavery@bgsu.edu

Joyce Nutta

University of Central Florida

Alison Youngblood

Western Kentucky University

Follow this and additional works at: https://scholarworks.bgsu.edu/seflp_pubs

How does access to this work benefit you? Let us know!

Repository Citation

Lavery, Matthew Ryan; Nutta, Joyce; and Youngblood, Alison, "Analyzing Student Learning Gains to Evaluate Differentiated Teacher Preparation for Fostering English Learners' Achievement in Linguistically Diverse Classrooms" (2018). *School of Educational Foundations, Leadership and Policy Faculty Publications*. 15.

https://scholarworks.bgsu.edu/seflp_pubs/15

This Article is brought to you for free and open access by the College of Education and Human Development at ScholarWorks@BGSU. It has been accepted for inclusion in School of Educational Foundations, Leadership and Policy Faculty Publications by an authorized administrator of ScholarWorks@BGSU.

Analyzing Student Learning Gains to Evaluate Differentiated Teacher Preparation for Fostering
English Learners' Achievement in Linguistically Diverse Classrooms

Matthew Ryan Lavery

Bowling Green State University

Joyce Nutta

University of Central Florida

Alison Youngblood

Western Kentucky University

Author Note

Matthew Ryan Lavery, School of Educational Foundations, Leadership and Policy, Bowling Green State University; Joyce Nutta, School of Teaching, Learning, and Leadership, University of Central Florida; Alison Youngblood, Department of English, Western Kentucky University.

Correspondence regarding this article should be addressed to Matthew Ryan Lavery, BGSU - Educational Foundations, Leadership & Policy (EFLP), 550 Education Building, Bowling Green, OH, 43403.

E-mail: mlavery@bgsu.edu

The Version of Record of this manuscript has been published and is available in *Journal of Teacher Education*, 07 Nov 2017, [doi:10.1177/0022487117751400](https://doi.org/10.1177/0022487117751400)

MATTHEW RYAN LAVERY is Assistant Professor of Classroom Assessment, Research, and Applied Statistics in Education for the School of Educational Foundations, Leadership, and Policy at Bowling Green State University. His research interests include multilevel analyses of classroom assessment data, Monte Carlo simulations, formative assessment, and instrument development and validation.

JOYCE NUTTA is Professor of English for Speakers of Other Languages Education in the School of Teaching, Learning and Leadership at the University of Central Florida. Her research focuses on preparing generalist teachers to meet the needs of English learners (ELs) in EL-integrated classrooms and the use of technology to teach second languages.

ALISON YOUNGBLOOD is an Assistant Professor of Teaching English to Speakers of Other Languages (TESOL) in the Department of English at Western Kentucky University. Her research interests include vocabulary development for English learners and TESOL curriculum development for pre-service teachers.

Abstract

Researchers compared pre/post classroom assessment scores of $n = 8326$ K-12 students taught by $n = 288$ teacher candidates to determine if a differentiated teacher education program prepared them to support English learners' (ELs) achievement in classrooms including native and non-native speakers of English. Candidates in Group 1 comprised academic subject (secondary mathematics, science, and social studies) teacher candidates, who completed six teacher preparation courses with 15 key assignments that included a focus on ELs. Certification areas for Group 2 candidates include language arts instruction (elementary, early childhood, and secondary English language arts). Group 2 candidates completed from 12-15 courses with from 41 to 50 key assignments that included a focus on ELs. Results indicate that teacher candidates in both groups helped narrow the gap between ELs and non-ELs from pretests to posttests. ELs performed no differently when taught by candidates from either group. Implications for teacher preparation are discussed.

Keywords: Preservice Teacher Education, Language Learners, TESOL, School/Teacher Effectiveness, HLM (Hierarchical Linear Modeling)

Analyzing Student Learning Gains to Evaluate Differentiated Teacher Preparation for Fostering English Learners' Achievement in Linguistically Diverse Classrooms

Teacher education programs face increasing demands in preparing candidates for 21st century classrooms. Many issues drive national discussions on teacher education, including using student learning data to inform instructional decision-making (Hamilton et al., 2009) and meeting rigorous standards for college and career readiness (Common Core State Standards Initiative, 2015). As teacher preparation programs strive to adhere to time-tested quality practices and face competition from alternate pathways to certification, they must also incorporate emerging content and practices by continuously looking ahead and preparing candidates for educational and demographic trends that cannot be ignored.

One trend is the growing number of English learners (ELs) in US PK-12 classrooms. There are over 4 million students needing English language support in the public schools. English learners make up 9% of children served in the classroom (National Center for Education Statistics, 2015c), and ELs are likely to have lower test scores and higher dropout rates (Ballantyne & Sanderman, 2008). In fact, at both elementary and secondary levels, around 70% of ELs score below basic proficiency in reading compared to about 20-30% of their native-speaking peers. ELs can close the gap somewhat in mathematics in elementary school; around 40% of ELs score below basic proficiency in mathematics compared to 15% of their peers. However, by the secondary level, 69% of ELs score below basic levels compared to only 26% of native speakers (NCES, 2015a, 2015b).

A Gap in Teacher Education

Teacher preparedness to address the needs of ELs is increasingly important because the No Child Left Behind Act and the Common Core State Standards have increased accountability

for ELs' academic achievement. However, the specific training that teachers receive to meet this accountability varies greatly. More than half of US states do not require English for speakers of other languages (ESOL) training for teacher candidates (Education Commission of the States, 2014; Samson & Collins, 2012; National Clearinghouse for English Language Acquisition, 2008). In states that do require ESOL, the content and quantity of training varies. Compiling information from the Education Commission of the States (ECS; 2014) and National Clearinghouse for English Language Acquisition (NCELA; 2008), only six states require all teachers to receive an ESOL endorsement or certification: Arizona, California, Florida, Massachusetts, New York, and Pennsylvania. Twelve additional states include coursework, training, or a reference to EL needs and best practices in their licensure requirements. Some university-based teacher education programs have developed minors in ESOL to address the need for more prepared teachers (Lucas, 2011), but requiring additional credits may not be feasible for many programs, just as adding a minor may not be possible for many teacher candidates. In sum, many children enter the classroom without proficiency in the language of instruction and many content area teachers enter the classroom feeling unprepared to teach language learners (McGraner & Saenz, 2009). This "classroom communication gap" (Nutta, Mokhtari, Strebler, Mihai, & Crevecoeur-Bryant, 2014, p. 2) has not been fully addressed in teacher education. In Florida, which has one of the largest EL populations, a court settlement outlined steps to close the classroom communication gap.

In the 1990s, the Florida Department of Education (FLDOE) reached a settlement with the League of United Latin American Citizens (LULAC) and affiliated EL advocacy groups regarding the rights of ELs to equal access to educational programs. The settlement, known as the Consent Decree (LULAC v. FLDOE, 1990), provided the framework to improve ELs'

education and set professional development requirements for teachers and administrators.

Teachers in academic subject areas like math, science, and social studies are required to complete 60 hours of professional development. Early childhood, elementary, English language arts, and exceptional education teachers are required to obtain an endorsement in ESOL, or the equivalent of 300 professional development hours (LULAC v. FLDOE, 1990).

When the Consent Decree was enacted, there was already an ESOL endorsement program available in Florida that included fifteen graduate credits in methods, curriculum and material development, assessment, cross cultural communication, and applied linguistics. These courses, however, were mostly offered in master's programs preparing university-level instructors of English as a second language or those interested in teaching English abroad, so they did not match the needs or background knowledge of certified K-12 teachers (Nutta, Mokhtari, & Strebel, 2012). There was also a strong push to integrate Consent Decree requirements into undergraduate teacher preparation curricula so that teachers met state mandates upon graduation. To further complicate matters, around the same time, Florida enacted a strict 120-hour limit on undergraduate degree programs. Florida's colleges of education were tasked with providing more training without increasing contact hours. These two incompatible mandates led to an innovation that came to be known as ESOL Infusion (FLDOE, 2011). Florida's ESOL Infusion model was developed as an alternative to adding 15 credits of ESOL coursework to some teacher education programs. Infused programs typically added six ESOL credits and integrated the content of the remaining nine credits into relevant existing coursework, such as methods of teaching reading courses.

Evolving Second Language Acquisition Theory

Beyond changing demographics and professional development requirements, theory and practice of how children learn English as a second language in schools also evolved. Thirty years ago, ELs received formal language instruction in an ESOL classroom while the mainstream classroom was viewed as an opportunity for natural language learning through exposure to the target language. This dichotomy made the content area classrooms a sink-or-swim environment as language learning had to precede content learning (Nutta, Mokhtari, & Strebel, 2012). This model is no longer tenable. ELs in today's classrooms are tasked with acquiring language and content proficiency simultaneously; therefore, ESOL teachers alone cannot help students be successful, and the underlying theory of second language acquisition (SLA) in public schools has evolved to reflect the important roles of content area and language teachers.

One of the current theories of SLA, the Interactionist perspective offers insights into the classroom environment requiring a collaborative approach to instruction. The Interactionist model emphasizes the role of communicative input that is comprehensible (Krashen, 1985), opportunities to produce meaningful language output (Swain, 1985), and encouraging students to interact verbally with fellow ELs, native speakers, and teachers (Long, 1996; Gass & Mackey, 2006). In the Interactionist model, these three elements are posited as essential to developing second language proficiency. While these elements appear to be commonplace in classrooms, the key for language learners is to provide opportunities for ELs to engage with them based on their current proficiency level. These opportunities encourage the simultaneous development of academic content and language knowledge. Providing these opportunities requires classroom teachers to understand SLA, to communicate with ELs so they comprehend the lesson topic, and to support ELs' communication so they can successfully demonstrate achievement of the

objectives. Concurrent with this natural, yet supported communication that occurs in the integrated classroom (ELs and non-ELs learning together), ELs benefit from ESOL instruction that is targeted to their stage of English language acquisition (Saunders, Goldenberg, & Marcelletti, 2013), or what is referred to as instructed SLA (Ellis, 2008). Where once the naturalistic second language learning environment of the integrated classroom operated without communication support and in isolation from ESOL instruction, now both ESOL and content area classrooms are integrated on a continuum of naturalistic and instructed SLA (Nutta, Mokhtari, & Strebel, 2012).

Closing the Classroom Communication Gap

Teacher education programs in Florida have infused ESOL teacher preparation into existing programs of study as a way of meeting state-mandated teacher training requirements of the Consent Decree and to address the classroom communication gap for teaching English learners. These programs include, at a minimum, added focus on ELs in existing teacher education courses and clinical experiences and one or more courses that focus exclusively on EL issues. While no claim can yet be made about the effect of the mandate on EL achievement, Samson and Collins (2012) analyzed 2009 National Assessment of Education Progress (NAEP) data regarding the percentage of students scoring at or above basic level in reading. The results indicate that among the top five states in EL enrollment, Florida has the narrowest gap between English learners and non-English learners: Florida ELs' scores were 22 points lower than non-ELs, with the gap for Texas at 33, California 41, Massachusetts 43, and New York 44. For the eighth grade, Florida's difference was 36 points, with California at 53, New York and Texas at 56, and Massachusetts at 59.

One specific example of an infused teacher education program in Florida is the One Plus Model of ESOL Infusion (Nutta, Mokhtari, & Strebel, 2012; Nutta, Mokhtari, Strebel, et al., 2014). The purpose of the present study is to explore the effectiveness of the One Plus Model at preparing two different groups of teachers to bridge the classroom communication gap. The following sections outline the curriculum and structure of the One Plus model and the rationale for separating teacher candidates into two groups. It is important to emphasize that the teacher preparation program in the current study uses the One Plus Model for all tracks.

The One Plus Model of ESOL Infusion

A major goal of the One Plus model is to foster an understanding of how teachers of different subjects and grade levels provide comprehensible instruction for ELs, enable successful communication between ELs and others, and support linguistic output of increasing complexity and accuracy. In addition, it emphasizes the divergent yet complementary roles teachers have in the academic success of ELs as outlined by the Consent Decree.

Categorization of Teacher Education Programs in the One Plus Model

The One Plus Model uses two groups. Teacher candidates in areas such as mathematics, science, and social studies are called Group 1. These teachers are responsible for supporting and developing EL's discipline-specific language and content knowledge, but are not responsible for general language and reading development. Elementary, English language arts, reading, exceptional and early childhood education programs emphasize the teachers' key role in promoting ELs' English language development in oral and literacy skills as well as content area benchmarks. These teacher candidates are called Group 2. These two levels of preparation for teaching ELs are derived from the Consent Decree mandate. The levels are also derived from common teacher preparation curricular practices, such as the typical inclusion of subject area

courses outside of teacher preparation programs for academic subject programs (whose curricular content are frequently outside of the purview of teacher education faculty and thereby more difficult to embed), as well as the rationale that language arts and literacy teachers need greater depth and breadth of preparation for teaching ELs to develop proficiency in English listening, speaking, reading, and writing.

Curriculum for Group 1 and Group 2

As noted in Table 1, the One Plus model curriculum prepares teacher candidates in Group 1 and Group 2 through ESOL-specific courses, ESOL-embedded content in respective core classes, and ESOL-related field experiences. The following paragraphs describe both groups' preparation in detail and are further summarized in Table 2.

The core curriculum. The core curriculum of math, science, and social studies in Group 1 and of elementary, English language arts, and early childhood in Group 2 (this institution currently does not offer undergraduate initial certification degrees in exceptional or reading education) have the same five courses in the core of their major: classroom management, ethical issues in education, learning theory and assessment, best practices for educating students with exceptionalities, and literacy practices.

Each core course contains ESOL-embedded content, instruction, and key assignments. For example, the core course on classroom management includes two ESOL-specific key assignments. All section instructors introduce candidates to three case studies representing an EL at beginning (“Edith”), intermediate (“Edgar”), and advanced (“Tasir”) proficiency. Candidates are then trained to identify language demands and contextual support in a lesson plan and supplement the lesson with verbal and nonverbal accommodations. This training is led by an ESOL Education faculty member or doctoral student as a guest lecturer. Candidates demonstrate

mastery of the material by adapting a lesson plan and suggesting specific accommodations appropriate for each case study (see Figure 1). The adapted lesson plan is followed by a microteach.

An additional core requirement is enrollment in an ESOL-specific course that focuses on the theory and practices of ESOL in K-12 classrooms. This course covers the historical context of ESOL in public schools, theories of first and second language learning, the multiple dimensions of 'knowing' a language, and developing skills in speaking, listening, reading, and writing a second language. While the course includes multiple assignments, three key assignments include an ESOL field experience, cultural analysis of a significant language minority group in the US, and an adapted lesson plan requiring more depth and scope of support for ELs. Key assignments in the core are completed by all candidates in all sections and are collected by Group 2 candidates specifically in an online assessment system that is reviewed by ESOL and course subject instructors and is linked to Florida Teacher Standards for ESOL Endorsement (2010).

Specialized infusion for Group 1. For Group 1 candidates, additional ESOL development comes through interacting with ELs during field experiences as mandated by their program and the completion of a Teacher Work Sample during student teaching. Note that some Group 1 candidates may not receive additional exposure to ELs based on their internship assignments, as linguistically diverse placements are not required for Group 1. The geographic area studied, however, suggests this would be the exception as ELs make up, on average, 10% of students in the five districts hosting most teacher candidates from this institution (the most common home language for ELs in this area is Spanish, followed by Haitian Creole; Florida

Department of Education, 2013). In addition, only students who worked with ELs during student teaching were included in the present study.

Specialized infusion for Group 2. Group 2 candidates have three additional unifying experiences in the One Plus Model. First, they all take a second course on applied linguistics which includes eight key assignments. Second, they are required to work with ELs during their internships and address ESOL best practices in all unit plans and reflections therein. Finally, each teacher candidate completes an online e-portfolio of graded key assignments from core and specialization courses. Each infused course syllabus designates which graded key assignment(s) need to be saved for the e-portfolio. The assignments are used to assess mastery of the state ESOL standards. The following paragraphs summarize the infused content in Group 2's specialization courses. Because the number of infused courses by program of study can be misleading, the actual number of key assignments in each course is listed in parenthesis (see also Table 2). Examples of key assignments from one infused course are also provided.

English language arts. The English language arts (ELA) program has infused five track specialization courses: history of the English language (2), survey of adolescent literature (4), literacy development in digital contexts (4), best practices in ELA (5), and teaching language and composition (3). The best practices in ELA course requires candidates to create a digital ELA resource library for ELs, write a research paper on the effectiveness of specific ESOL teaching strategies, complete an observation of an ESOL classroom, reflect on their own experiences as a language learner, and design a unit plan of instruction with ESOL accommodations and assessment strategies for the three case studies introduced in the core.

Elementary education. The elementary education program has infused eight specialization courses: science (3), math (2), language arts (3), social studies (2), and art (1) for

elementary school children as well as courses on foundations of reading development (2), children's literature (4), and diagnosis and correction of problems in literacy development (2). The infused assignments in the language arts for elementary education course require candidates to collect model writing samples for native and non-native speakers in different rhetorical modes, to assemble a set of classroom activities each with ESOL accommodations for three proficiency levels, and to compose a letter home to parents with versions that are accessible for non-English speaking households.

Early childhood education. The early childhood education program has infused nine of their specialization courses: early childhood literacy (3), mathematics (4), science and technology (3), parental involvement (4), exceptional education (6), activity development (2), curriculum and planning (4), assessment (3), and a bridge course to prepare students for their internship (2). In the assessment course, key assignments have students evaluate an assessment for young ELs, observe the assessment of an EL, and write a paper on fair and ethical assessment for young ELs.

Instructional Fidelity and Quality of Infused Content

A final note on the specific ESOL Infusion curriculum tested in this study relates to the quality of the instruction and key assignments in the core and specialization courses. Faculty who teach these courses have varying levels of expertise and experience with ESOL as a discipline. To promote quality and continuity of instruction between the programs and even among multiple sections of the same course, an ESOL Education faculty member or doctoral student was assigned to each program. The ESOL specialist worked directly with individual program faculty and lead course instructors to infuse authentic ESOL elements into existing assignments, generate new assignments as needed, and provide ESOL-specific course readings

and activities that matched the course learning objectives. The ESOL specialists also collaborated to ensure the infusion was comprehensive by creating a curriculum map of when and how each state ESOL standard was addressed and assessed throughout each program. The use of standardized assignments and pre-selected materials was a critical step in implementing the One Plus model.

Research Question

In the context of education policy, some initiatives align equal preparation of teachers with equal opportunity for students (Harper, de Jong, & Platt, 2008). In the case of ESOL, this approach would require all teachers to receive 300 dedicated, as opposed to infused, hours of ESOL. By exploring the impact of teacher candidates from Groups 1 and 2 on English learners' achievement, the premise of the stratified One Plus model was examined, through the following research question:

Are English Learner unit post-assessment scores related to whether the teacher candidate is enrolled in a Group 1 or Group 2 ESOL-infused teacher preparation program, after controlling for student pre-assessment score, disability status, minority status, and qualification for free / reduced price lunch?

Method of Inquiry

Teacher candidates in the current study completed the One Plus Model of ESOL Infusion. The present study examines the One Plus model by investigating the pre and post assessment scores of English Learners (ELs), compared to native speakers of English and students who have exited from ESOL programs (non-ELs), taught by two groups of teacher candidates during student teaching: 1) teacher candidates enrolled in Mathematics Education, Science Education, and Social Science Education programs (*Group 1*), and 2) teacher candidates enrolled in

Elementary Education, Early Childhood Education, and English Language Arts Education programs (*Group 2*).

Data Source and Instruments

The teacher education institution studied uses the Teacher Work Sample (TWS) during student teaching placements. The TWS is reflective documentation of an instructional unit taught by the teacher candidate who uses pre and post assessments to measure student learning gains (Henning, Kohler, Robinson, & Wilson, 2009; Rosselli, Girod, & Brodsky, 2011). At the institution studied, the TWS is a standardized assignment for which all teacher candidates receive the same instructions and training, and are scored according to the same rubric. The TWS is a key assignment in all initial teacher preparation programs. While teacher candidates are encouraged to collaborate with the classroom teacher who hosts their internship in all other aspects of student teaching, they are expected to plan, assess, and teach the TWS unit independently. Though it is safe to assume that there is some variance across placements in the degree of autonomy granted to teacher candidates, cooperating teachers are all notified that the TWS unit is intended to assess the teacher candidate's ability to design and deliver an entire instructional unit on their own.

Teacher Work Sample data were collected using the TWS GraphMaker (Version 5.1.2; Lavery, 2012), a software tool that generates multiple graphs of student pre and post assessment scores to assist teacher candidates in analyzing student learning. The TWS GraphMaker also allowed the institution to compile anonymized pre/post assessment data and demographic information for K-12 students taught by the teacher candidates exposed to the One Plus model. The data set used in the present study is comprised of TWS data submitted by teacher candidates who completed a Teacher Work Sample for one of the academic education programs listed prior.

By systematically examining the pre and post assessment data, researchers compared the scores of ELs and non-ELs taught by teacher candidates in Groups 1 and 2.

As part of TWS training, teacher candidates are instructed to use either (a) identical instruments, (b) identical subsets of items from similar instruments, or (c) similar performance tasks scored using identical rubrics (e.g., essays written for similar prompts scored with the same rubric) for the pre and post assessments. The TWS GraphMaker requires the scores for pre and post assessments to be entered with a common structure and format (i.e., same number of available points for each standard or learning goal assessed). Though it is beyond the scope of the present study to systematically analyze the content of submitted TWS artifacts (see Limitations and Future Research section), researchers' experience at this institution suggests that the vast majority of teacher candidates follow recommendation (a), most of the remaining teacher candidates follow recommendation (b), and very few teacher candidates attempt recommendation (c). Since only a fraction of the teacher candidates who attempt (c) are expected to produce non-parallel assessments, researchers assume that the pre and post assessment measures included in the present study produce comparable scores.

Researchers also assume that systematic measurement error inherent in each assessment instrument will affect student's pre and post assessment score similarly. By including pre-assessment scores in analyses of post-assessment scores, measurement error associated with the assessment is represented on both sides of the equation. Thus, this source of error subtracts itself out, preserving the variance associated with student academic growth. The present study takes advantage of a common practice in the physical sciences; when the precise calibration (i.e., the accuracy of zero on the scale) of an instrument cannot be known with confidence, initial and final measurements may be taken with the same instrument to use the difference between scores

to remove instrumental error (see Dartmouth College, 2011). Teacher candidates' classroom assessments can be expected to vary in overall assessment quality, and can be expected to have greater variance in measurement error than professionally produced, standardized instruments. Researchers further assume that measurement error is approximately normally distributed with a mean of zero across all measurement instruments. Although measurement error is expected to cause systematic error within classrooms, such error is expected to be uncorrelated across classrooms, thereby reducing the magnitude of effects and increasing the need for statistical power. Variations in assessment quality are not expected to skew or distort effects given a large enough sample, making TWS assessments suitable for the present study.

Classroom Assessments

Though classroom assessment scores may not seem as robust a measure as standardized tests, it allows evaluation of key aspects of candidates' performance with English learners. One principle emphasized in the One Plus Model is the importance of congruence among differentiation strategies for ELs across curriculum, instruction, and assessment. This means if they teach a lesson on the water cycle, for example, they will use the same visuals used during instruction to enable a beginning English learner to label the processes during a unit test. If differentiation strategies are used consistently across curriculum, instruction, and assessment, then we would expect classroom-based pre- and post-assessments to measure learning gains more accurately for English learners than a standardized assessment which may use unfamiliar vocabulary and presentation to assess content which the EL understands. While a native English speaker may not be challenged by multiple linguistic representations of familiar material, a single unfamiliar vocabulary word or non-standard expression may introduce measurement error for ELs. Classroom-based tests are also an important factor in determining grades, so examining

to what degree candidates are successful in making the content of their classroom assessments accessible to ELs could indicate candidates' effectiveness in providing equitable education for ELs. Finally, classroom assessments tend to be more directly aligned with an individual teacher's instructional objectives for students, and are therefore expected to measure the intended outcomes of that instruction more accurately. Although prior research on the academic achievement of ELs relies almost exclusively on standardized and large-scale instruments, classroom assessments may be better suited to the present research question.

A Priori Power Analysis

To determine whether the present research design possesses sufficient statistical power to answer the proposed research question, the researchers analyzed multiple data sets constructed to resemble observed data in a Monte Carlo simulation study design (Fan, Felsevalyi, Sivo, & Keenan, 2002; Vogt, Gardner, & Haeffele, 2012). The simulated data set contained teacher candidates from both groups analyzed, including randomly generated samples of students with attributes that reflect observed descriptive statistics for students in the present study. Student scores on teacher candidate assessments were generated such that the classroom-mean measurement error for each teacher candidate varied (to represent easy tests, difficult tests, accurate tests, and less reliable tests), while keeping the grand-mean measurement error equal to zero. Two simulated conditions were examined; one in which each group of teacher candidates taught ELs equally well (or equally poorly), and one in which known interaction effects were introduced into the data of varying magnitudes (representing that one group fostered higher learning gains for ELs than the other group). The magnitude of the EL effect on test scores was systematically varied to test the sensitivity of the present research design to detect effects of different sizes. Finally, to test statistical power under different sample sizes, both conditions were

simulated for a sample of $n = 30$ teacher candidates in each group (for a total of $n = 60$ teacher candidates), and for $n = 100$ teacher candidates in each group ($n = 200$ total).

Though a detailed discussion of the Monte Carlo simulation study is beyond the scope of the present paper (see Lavery, 2017), the findings of the simulation suggest that the models analyzed (described in the next section) are sufficiently powerful to answer the present research question with confidence. Figure 2 presents paired histograms arranged in a population pyramid style. The panels on the left indicate simulation runs in which the parameter of interest (either EL or the interaction effect) was non-significant, while the panels on the right indicate that the parameter of interest was significant. The simulation study suggests that, when the academic disadvantage attributable to EL status exceeds one percentage point for a small sample ($d = 0.05$ under observed conditions, a very small effect; Cohen, 1988), or exceeds half a percentage point for a larger sample ($d = 0.03$, a very small effect), the present research design will accurately detect the difference. The simulation further indicates that, when the difference between the mean post-assessment scores of ELs taught by one group of teacher candidates and the scores of ELs taught by the other group exceeds half a percentage point for a small sample ($d = 0.03$), or exceeds 0.4 percentage points with a larger sample ($d = 0.02$), the present research design will accurately detect it.

A second simulation designed to replicate the conditions observed (i.e., with $n = 288$ teacher candidates with class sizes and student demographics which reflect the observed data) confirmed the statistical power of the first simulation but indicated an inflated Type I error rate. To maintain a 5% Type I error rate across 10,000 replications of the simulation, researchers must use $\alpha = .019$ for level-one effects (i.e., student characteristics such as *EL*), $\alpha = .051$ for level-two effects (i.e., teacher candidate characteristics such as *Group*), and $\alpha = .011$ for cross-level

interactions (such as *Group*EL*). Thus, since the analysis of teacher made pre/post assessments demonstrates adequate statistical power for the present study, researchers proceeded with analysis of TWS unit assessments using the alpha levels determined via simulation.

Hypothesized Model

Variance in student post-assessment scores is likely to stem from several sources. Using a covariate adjustment model like those used in value-added analyses (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004), the present study adjusts its estimates of student learning by including available covariates that have been theoretically and empirically linked to student learning gains. A two-level model shown in equation (1), in which second level units are teacher candidates and first-level units are the K-12 students taught, assessed the effects of student pre-assessment scores (*Pre*), English Learner status (*EL*), disability status (*SWD*), minority status (*Minority*), whether the student qualifies for free or reduced price lunch programs (*LowSES*), as well as whether the teacher candidate is enrolled in a Group 1 or Group 2 ESOL infusion program (*Group*, dummy coded such that *Group 2 = 1*), on the student's post-assessment score (*Post*). Multilevel analyses were conducted with PROC MIXED in the SAS 9.4 software package using restricted maximum likelihood estimation (Raudenbush & Bryk, 1996; Singer, 1998; Snijders & Bosker, 1999).

$$Post_{ij} = \gamma_{00} + \beta_{1j}Pre_{ij} + \beta_{2j}Minority_{ij} + \beta_{30}EL_{ij} + \beta_{40}SWD_{1j} + \beta_{5j}LowSES_{ij} + \gamma_{01}Group_j + \beta_{6j}EL_{ij}Group_j + (u_{0j} + r_{ij}) \quad (1)$$

In the notation of equation (1), β s are used to represent fixed parameters directly estimated using all level-one units regardless of classroom and γ s are used for effects estimated at level-two or parameters estimated separately for each classroom and thus allowed to vary across teacher candidates. The hypothesized model adjusts parameter estimates by controlling for prior achievement ($\beta_{1j}Pre_{ij}$). Equation (1) includes random intercepts (γ_{00}), which reflect the

common variance in student test scores that can be attributed to the classroom or teacher level, but which cannot be explained by any of the other variables in the model. Such variance is sometimes attributed to teacher effectiveness and is often called the teacher effect in value-added and teacher evaluation literature (Konstantopoulos, 2014; Kupermintz, 2003; McCaffrey et al., 2004). Note that the output of PROC MIXED reports random effects as both a fixed effect estimated across all clusters (or the grand mean effect), and a random effect estimated for each cluster (the degree to which the classroom mean deviates from the grand mean). Thus, the γ_{00} term in equation (1) represents both the grand mean score observed across clusters (β_{00}), as well as the degree to which the mean post-assessment score in each classroom deviates from the grand mean (γ_{00}).

The remaining parameters in equation (1) represent the degree to which students in each dummy-coded subgroup differ from the mean performance of their peers. Controlling for prior knowledge of the content taught ($\beta_{1j}Pre_{ij}$), and removing the variance associated with the grand mean post-assessment score (β_{00}) as well as the classroom specific mean post-assessment score (γ_{00}), allows interpretation of the other parameters in the model as differential learning gains. The hypothesized model contains the interaction term $\beta_{6j}EL_{ij}Group_j$, reflecting that teacher candidates from Group 2, who have more extensive training to foster English language development may be expected to differ in their abilities to foster learning gains for ELs and other students. All variables except for pre-assessment score are dummy-coded to indicate group membership. Thus, the parameter β_{6j} represents the differential learning gains observed for ELs taught by teacher candidates in Group 2, compared to all other students. As robustness tests, researchers also analyzed the full final data set described next using normal curve equivalence (NCE; Gamel, Tallmadge, Wood, & Binkley, 1974; Wright, White, Sanders, & Rivers, 2010; Yen

& Candell, 1991) scores, and also analyzed randomly selected subsets of the final data set. Both additional analyses are explained in more detail after discussion of the findings of the first analysis.

Data Set and Sample

The initial data set was comprised of data for $n = 14,833$ K-12 students reported by the $n = 498$ teacher candidates who completed a Teacher Work Sample during a single year. This study focused on the five major disciplines of initial teacher certification, namely elementary education, English language arts education, mathematics education, science education, and social science education (teacher candidates from other areas, such as French education or art education, were dropped). Since the present study seeks to compare ELs and non-ELs, teacher candidates who reported data for only one of these subgroups of students were removed from the data set. Teacher candidates who did not report ESOL status for any of their students were also dropped from the analysis since the researchers could not be certain that they taught both ELs and non-ELs. Individual student records with missing data for ESOL status were retained for analysis with multiple imputation (described next). The final data set contained anonymized data with pre and post-assessment scores and demographic information for $n = 8,326$ K-12 students and $n = 288$ teacher candidates. Each teacher candidate reported data on from 10 to 145 students ($M = 28.9$, $SD = 24.0$). The final data set includes K-12 students from a variety of diverse settings including urban, suburban, and rural schools; elementary, middle, and high schools; and from both large and small schools in up to 11 counties.

The final data set was examined for missingness. No level-two variables contained missing data for any of the teacher candidates in the analysis. The number and percent of student records with missing data is displayed for each level-one variable in Table 3. All level one

variables in the analysis contain less than 1% missing data, except for *LowSES*, which is 15% missing. Per researchers' experience, one of the districts at which teacher candidates are placed considers students' free or reduced-price lunch status (on which the *LowSES* variable is based) as protected information and does not share that data with teachers. Although missingness on this variable is assumed to be systematically related to the district at which the teacher candidate completed the TWS, the district in question is not systematically different from the other districts served by this institution in the prevalence of ELs or in district ESOL policies. Missingness on the other variables is not expected to relate to post-assessment scores after controlling for observed variables. Thus, researchers assumed all variables to be missing at random and suitable for multiple imputation (Little & Rubin, 2002). Researchers used Blimp (a software tool designed for multiple imputation of multilevel data; Enders, Keller, & Levy, 2017, Keller & Enders, 2017) to generate 25 multiply imputed datasets for analysis with PROC MIXED as described prior. Results from the multiple analyses were then combined using PROC MIANALYZE. Also note that while the findings reported in the next section use multiply imputed data, imputations were not used to calculate descriptive statistics reported in this and prior sections.

Table 4 displays the mean number of students in each cluster reported by teacher candidates in both groups by subgroup of student. As teacher candidates were only included in the present study if they reported data on both ELs and non-ELs, every cluster contains both subgroups of students. Across all clusters, teacher candidates reported data on from one to 56 ELs ($M = 4.3$, $SD = 6.0$) representing from 1.1% to 94.4% of the students they taught ($M = 15.9\%$, $SD = 12.7\%$). Researchers did not select teacher candidates for inclusion in the study based on the proportions of other student subgroups reported. As such, 56 teacher candidates did not report data on students with disabilities ($n = 5/56$, or 9%, of *Group 1* and

$n = 51/232$, or 22%, of *Group 2*). Likewise, 9 teacher candidates, all in *Group 2* (4% of the group), did not report data on students who qualify for free or reduced-price lunch, though 25 teacher candidates ($n = 1/56$, or 2%, of *Group 1* and $n = 24/232$, or 10%, of *Group 2*) reported 100% of their students as *LowSES*. Finally, although all teacher candidates reported data on at least one minority student, 10 teacher candidates ($n = 1/56$, or 2%, of *Group 1* and $n = 9/232$, or 4%, of *Group 2*) identified all students as minority students. Although 81 teacher candidates ($n = 7/56$, or 13%, of *Group 1* and $n = 74/232$, or 32%, of *Group 2*) retained in the sample did not teach both members and non-members of one or more dummy-coded demographic group, researchers chose not to limit the sample further by removing them. Since only teacher candidates who taught both ELs and non-ELs were included in the analysis, the comparison of interest is preserved across all clusters analyzed and no further reduction of statistical power is necessary.

Findings

Consistent with strong, well-documented, historical trends seen in standardized testing (NCES, 2015a, 2015b), ELs in this data set demonstrated lower classroom pre-assessment scores ($M = 36.1$, $SD = 23.2$) than non-ELs ($M = 41.7$, $SD = 24.7$), $t(8324) = -7.48$, $p < .001$, $d = -0.23$, as well as lower post-assessment scores ($M = 69.7$, $SD = 21.6$) than non-ELs ($M = 75.4$, $SD = 20.5$), $t(8324) = -9.06$, $p < .001$, $d = -0.28$. The observed effects are both small (Cohen, 1988). Even when pre-assessment scores are included as covariates, differences between marginal means estimated at $Pre = 40.8$ for EL post-assessment scores ($M = 71.3$, $SD = 18.8$) and non-EL scores ($M = 75.1$, $SD = 18.8$) are significant, $F(2, 8323) = 895.97$, $p < .001$, though still small, $d = -0.20$. Researchers analyzed the TWS data with the full model shown in equation (1) to understand the nature of these differences.

Full Multilevel Model

The $\beta_{6j}EL_{ij}Group_j$ term did not significantly explain post-assessment scores, $t(8156.4) = -0.74, p = .457$, indicating that ELs taught by Group 2 Language Arts or Elementary teacher candidates performed no differently than their non-EL classmates, and no differently than ELs taught by Group 1 teacher candidates on TWS unit assessments. Researchers dropped the non-significant interaction, evaluating relative model fit with the likelihood ratio test (Meng & Rubin, 1992). The hypothesized model shown in equation (1) did not fit the data better than the model without the non-significant interaction term, $F(1, 895282.337) = 2.73, p = .099$. Thus, we report the results of the final model,

$$Post_{ij} = \gamma_{00} + \beta_{1j}Pre_{ij} + \beta_{2j}Minority_{ij} + \beta_{30}EL_{ij} + \beta_{40}SWD_{IJ} + \beta_{5j}LowSES_{ij} + \gamma_{01}Group_j + (u_{0j} + r_{ij}) \quad (2)$$

Estimates produced by the final model are shown in Table 5. All analyzed level-one variables significantly explain student post-assessment scores at the chosen alpha level of $\alpha = .019$. The level-two parameter $\gamma_{01}Group_j$ was non-significant, $t(8290.2) = 1.65, p = .098$, indicating that student scores were not related to teacher candidate group, after controlling for the other variables in the model.

Normal Curve Equivalence (NCE) Scores

Teacher candidates in this data set did not teach the same content, teach in the same context, administer the same test (or even comparable tests) as other teacher candidates. Researchers assume that the pre and post assessment administered within the same cluster are parallel tests producing comparable scores, but the same cannot be said when making comparisons across clusters. To determine whether equation (1) adequately removes construct irrelevant variance to address the research question, researchers analyzed the data set using normal curve equivalence (NCE; Gamel et al., 1974; Wright et al., 2010; Yen & Candell, 1991)

scores. NCE scores are z-scores transformed to resemble percentile ranks (Gamel et al., 1974; Wright et al., 2010), but possess an equal interval scale and avoid floor and ceiling effects. Wright et al. (2010) report that the SAS[®] EVAAS[®] uses NCE scores when equivalent or vertically equated tests are not available across years. Since each teacher candidate used different tests, differences between assessment instruments may confound the analyses presented prior. Researchers used NCE scores to test for robustness.

Researchers used classroom-specific z-scores to generate NCEs, group-mean centering the pre and post assessment scores. This approach removes all classroom-specific variance from either differences in assessments, differences in quality of instruction, and differences in classroom specific contexts (for example, the influence of particularly skilled mentor teachers). Group-mean centered NCE scores preserve variance that lay between subgroups, however, supporting meaningful comparisons across classrooms. NCE scores were analyzed with a non-nested model. As with the initial model, the interaction term was non-significant, $t(8251.6) = -1.84, p = .065$, and was dropped, resulting in

$$\begin{aligned} PostNCE_i = & \beta_0 + \beta_1 PreNCE_i + \beta_2 Minority_i + \beta_3 EL_i + \beta_4 SWD_i + \beta_5 LowSES_i \\ & + \beta_6 Group_j + r_i \end{aligned} \quad (3)$$

All variables significantly explained post-assessment NCE scores except for $\beta_6 Group_j$, $t(7846.8) = 0.83, p = .408$, which is expected since classroom-specific variance was removed. Fixed effects for equation (3) are shown in Table 6. ELs scored lower on post-assessments than non-ELs, $\beta_3 EL_i = -2.10, t(8009.3) = -3.43, p = .001$. NCE scores are also ideal to establish whether ELs performed differently than non-ELs on pre-assessments (i.e., before receiving the relevant instruction), after controlling for the other variables in the model. Table 7 displays the results of an analysis using the equation,

$$PreNCE_i = \beta_0 + \beta_2 Minority_i + \beta_3 EL_i + \beta_4 SWD_i + \beta_5 LowSES_i + \beta_6 Group_j + \beta_7 EL_i Group_j + r_i, \quad (4)$$

EL status significantly explained pre-assessment scores, $\beta_3 EL_i = -5.31$, $t(8118.5) = -7.99$, $p < .001$, suggesting that ELs began the instructional units analyzed with less prior knowledge than their non-EL peers. Teacher candidate group remained non-significant, $t(7468.4) = 1.40$, $p = .163$, suggesting that each group of teacher candidates studied were assigned students with no differences in mean prior knowledge of the content taught.

The parameter estimates reported in Table 6 and Table 7 indicate the mean difference in classroom assessment scores on the content taught for students in each of the dummy-coded groups, after controlling for the other variables in the analysis. Thus, after controlling for the variance explained by the other variables included in equation (4), ELs scored 5.3 NCE points lower ($SE = 0.67$) on classroom pre-assessments than their non-EL peers, $d = -0.26$ (which is a small effect, Cohen, 1988). EL status had a greater effect on student pre-assessment scores than socioeconomic status ($\beta_5 = -3.3$, $SE = 0.52$) or minority status ($\beta_8 = -1.9$, $SE = 0.50$), and slightly smaller than disability status ($\beta_4 = -5.8$, $SE = 0.67$). The effect of EL status fell to $\beta_3 = -2.1$ ($SE = 0.61$) on post-assessment NCE scores, $d = -0.10$ (which is a very small effect). EL status had a similar effect on post-assessment scores as socioeconomic status ($\beta_5 = -2.1$, $SE = 0.48$), a larger effect than minority status ($\beta_8 = -1.2$, $SE = 0.46$), and was associated with less than half of the difference in NCE post-assessment scores as disability status ($\beta_4 = -5.7$, $SE = 0.62$). The NCE analyses suggest that the achievement gap narrowed for ELs between classroom pre-assessments and post-assessments, and that observed differences in student achievement after instruction are better explained by disability and socioeconomic than by EL status.

Randomly Selected Sub-Samples

Since a majority of the teacher candidates in this study are members of *Group 2* ($n = 232/288, 81\%$), unequal sample sizes may distort the findings. As a robustness test, researchers analyzed randomly selected, equal-sized sub-samples of *Group 1* and *Group 2* teacher candidates. A SAS macro selected 30 teacher candidates from each group and analyzed the final multilevel model and the NCE score model over 1000 replications, testing for the significance of each variable analyzed (using alpha levels from previous analyses). Each analysis included $\bar{n} = 2399.9$ ($SD = 141.6$) students, of which $\bar{n} = 317.7$ ($SD = 43.9$) were ELs and $\bar{n} = 660.0$ ($SD = 77.8$) were taught by teacher candidates in *Group 2*. Results of the analyses are shown in Table 8. Across all runs, the sign and magnitude of parameter estimates was similar to those found previously (compare to Tables 4 and 5). Notably, the mean effect size of EL status was smaller than that of any other student characteristic analyzed. These results suggest that differences in student post-assessment scores are better explained by student socioeconomic status, by disability status, and to a somewhat lesser degree, by minority status than by EL status.

Discussion

The fundamental goal of the One Plus ESOL Infusion curriculum model is the elimination of significant achievement differences between English learners and native speakers in the K-12 classroom by differentiating the preparation teacher candidates receive. This preparation trains teachers to adequately accommodate the educational and linguistic needs of ELs across content areas. As researchers, we are encouraged by these findings that suggest the ESOL Infusion curriculum might be better than the status quo at closing the classroom communication gap, as teacher candidates prepared through the One Plus model narrow the achievement gap between ELs and non-ELs. Further, no difference is observed between the

abilities of Group 1 and Group 2 teacher candidates to foster learning gains for ELs, suggesting that the One Plus model provides teacher candidates in various programs with instruction in second language acquisition and English language development commensurate with the role they play in helping ELs master the language demands of their respective content areas. In short, the One Plus model trains teacher candidates to be interactionists: they can generate comprehensible input (Krashen, 1985), provide opportunities for ELs to produce meaningful language output (Swain, 1985), and encourage ELs to interact verbally with fellow ELs, native speakers, and teachers (Long, 1996; Gass & Mackey, 2006). Most importantly, the One Plus model gives teacher candidates the tools to provide access points to these three elements at various levels of English proficiency.

The current study may help bridge promising research on in-service training approaches for teachers in STEM fields to use scaffolded linguistic and content support to close the classroom performance gap (Lara-Alecia, Tong, Irby, Guerrero, Huerta, & Fan, 2012; Llosa et al., 2015; Santau, Maerten-Rivera, & Huggins, 2011). For example, Santau et al. (2011) researched the ability of fourth grade science teachers ($n = 55$) to close the gap between EL and non-EL students on researcher-generated pre and post-tests after participating in in-service trainings. The trainings outlined the district's science curriculum with specific EL strategies, accommodations, and scaffolded language support. These targeted STEM examples also focus on understanding SLA, communicating with ELs about the lesson topic, and supporting ELs' linguistic development so they can successfully demonstrate achievement of the academic objectives. In the same spirit, the One Plus Model aims to shift the development of these ESL teaching abilities to the pre-service level for future generations of educators and create space

during in-service training for more advanced exploration and innovation of linguistically diverse classrooms.

Additionally, we intend for this study to add to the larger debate of exactly how much preparation all teachers should have to reach ELs. Although current initiatives encourage secondary academic subject teachers to incorporate language and literacy activities in subject area instruction (Common Core State Standards Initiative, 2015), they are not the major focus of instruction as they are when teaching, for example, written composition or developmental reading. While literacy scholars agree that academic subject teachers need to know about and do more to support their students' disciplinary literacy (International Literacy Association, 2015), there is no clear consensus about the amount of preparation that they need in literacy education. The same holds true for the preparation of all teachers to support English language development (ELD) for ELs. If all teachers are reading teachers and all teachers are ELD teachers, certainly some are to a greater degree than others. If those whose roles and responsibilities are primarily to teach academic subjects are expected to teach ELD in the same way and in the same amount as teachers whose roles focus on teaching language arts, do we know that the content of this preparation is directly relevant to the knowledge and skills that these teacher candidates have developed and will use in the classroom, and do we know that the time and money invested is well spent? Few would argue that less education about any important area is preferable, but when does more education become enough? The One Plus model posits the need for a greater degree of knowledge and skills in language theory and pedagogy for the primary goal of teaching ELs oral proficiency and literacy skills in English. This approach also differs from a simple addition of the same sequence of entirely EL-focused courses, as with a minor in ESOL, which would be attached to degree programs leading to certification. In contrast, the One Plus model mirrors the

future teacher's experience by showing EL education and assessment alongside native speakers in infused programmatic courses while providing a space for deeper development in EL-specific courses.

Limitations and Future Research

The findings of this study are limited by the use of teacher candidate generated and reported pre and post assessment scores. Thus, the validity and reliability of these teacher candidate-made assessments cannot be known. Some of the variability in student scores may be attributable to variation in the quality of these assessments, rather than to variability in the quality of instruction delivered by teacher candidates. Another factor is the subject and topic of the assessments, which were not identified in the data collected from the teacher work sample. For elementary teacher candidates, any subject could be the focus of the assessment since elementary faculty teach other academic subjects as well as English language arts. In contrast, secondary teachers would only be expected to assess their respective subjects. If a disproportionate number of the elementary candidates assessed mathematics, for example, their specific preparation in language arts development may not be put to the test. Future research will include the content domain of the lessons taught in order to examine these differences more closely.

Further, the nature of the data set used in the present study precluded the researchers from including additional, potentially informative covariates, such as dispositions or measures of related knowledge and skills. Because the data set had been deidentified prior to the present study, researchers could only analyze the variables already contained within; it cannot be known whether the two groups were equivalent prior to treatment. A large portion of the teacher candidates prepared by the institution studied are transfer students, preventing accurate records

about the prior academic preparation (e.g., high school GPA, ACT or SAT scores) of teacher candidates in the programs examined. It is possible that weaker teacher candidates are selected in to programs which receive stronger preparation, masking differences in the training groups. In addition, all the teacher candidates studied were exposed to the One Plus model, so research is needed to compare this model to other approaches of ESOL training and to programs that require no ESOL training at all. Thus, although we are able to examine the differences between student groups (i.e., ELs and non-ELs, students with and without a disability, etc.) before and after instruction, the present design does not allow estimation of the effects associated with these student characteristics when taught by teacher candidates who received different preparation, or no preparation at all, to teach ELs. Also, since this research uses classroom assessments, rather than standardized assessments, these findings cannot be compared to those of other studies until further research is conducted in a variety of contexts using aggregated classroom assessment results.

Future research will incorporate potentially informative covariates that were not available in the current data set. For example, it is possible that the age or grade-level of the students explains a portion of the variance in learning gains. While the present data set does not include student grade-level, similar data currently being collected at this institution for future analysis includes this variable. A dichotomous variable indicating elementary and secondary-level students was evaluated for inclusion in the present study, but did not add to the model and no further data was collected on its parameter estimates. It is also possible that the mentor teachers with whom the teacher candidates are placed may impact the findings. The use of anonymized data prevented linking teacher candidates to mentor teachers, and this may also be examined in future research. Finally, the present study is also limited by the fact that the specific English

language proficiency levels of the ELs taught are not known. In a sample as large as the one used in this study, one can assume relatively equivalent distributions of EL proficiency levels across groups, but without having that data to include in the analyses, the specific effect cannot be known. Future research is needed to probe these specific limitations and to investigate these phenomena further and additional data is currently being collected to support such studies.

Conclusion

The No Child Left Behind Act and the Common Core State Standards increased the expectations of teachers' ability to reach ELs to develop content and linguistic knowledge. In an effort to meet the demand for highly qualified teachers ESOL training is structured into a one-size-fits all approach, both teachers and ELs are left at a disadvantage (Nutta, Mokhtari, & Strebel, 2012; Harper, de Jong, & Platt, 2008; Langman, 2003). The findings of this study suggest that a multi-leveled EL-infusion preparation program can adequately prepare teacher candidates to deliver course content in order for ELs to make gains that take into account the demand of growing both language and content knowledge, and consequently, may hold promise for narrowing the performance gap between ELs and native speakers in the K-12 classroom.

References

- Ballantyne, K.G., Sanderman, A.R., Levy, J. (2008). *Educating English language learners: Building teacher capacity*. Washington, DC: National Clearinghouse for English Language Acquisition. Retrieved from <http://www.ncela.us/publications>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: L. Erlbaum Associates.
- Common Core State Standards Initiative (2015). *English language arts standards*. Retrieved from <http://www.corestandards.org/ELA-Literacy/>.
- Dartmouth College. (2011). *Computing uncertainties in laboratory data and result*. Retrieved from <https://www.dartmouth.edu/~chemlab/info/resources/uncertain.html>
- Education Commission of the States (2014). *What ELL training, if any, is required of general classroom teachers?* Retrieved from <http://ecs.force.com/mbdata/mbquestNB2?rep=ELL1415>
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.
- Enders, C. K., Keller, B. T., & Levy, R. (2017). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, No Pagination Specified. Advance online publication. doi:10.1037/met0000148
- Fan, X., Felsovalyi, A., Sivo, S. A., & Keenan, S. C. (2002). *SAS® for Monte Carlo studies: A guide for quantitative researchers*. Cary, NC: SAS Institute.
- Florida Department of Education. (2011). *Technical assistant for teacher preparation: Meeting the needs of English language learners (ELL) in Florida*. Retrieved from <http://www.fldoe.org/core/fileparse.php/7719/urlt/0071749-mnellf.pdf>

Florida Department of Education. (2013). *English language learners*. (Series 2013-09D).

Tallahassee, FL: Education Information & Accountability Services Data Report.

Florida Teacher Standards for ESOL Endorsement (2010). Retrieved from

http://education.fiu.edu/docs/esol_standards.pdf

Gamel, N. N., Tallmadge, G. K., Wood, C. T., & Binkley, J. L. (1975). *State ESEA Title I reports:*

Review and analysis of past reports, and development of a model reporting system and format. Retrieved from <http://eric.ed.gov/?id=ED118684>

Gass S. & Mackey, C. (2006). Input, interaction, and output: An overview. *AILA Review*, 19, 3-17. doi:10.1075/aila.19.03

Hamilton, L., Halverson, R., Jackson, S.S., Mandinach, E., Supovitz, J.A., & Wayman, J.C,

(2009). *Using student achievement data to support instructional decision making. IES practice guide*. NCEE 2009-4067. Retrieved from <http://eric.ed.gov/?id=ED506645>

Harper, C., de Jong, E., & Platt, E. (2008). Marginalizing English as a second language teacher expertise: The exclusionary consequence of No Child Left Behind. *Language Policy*, 7, 267-284. doi:10.1007/s10993-008-9102-y

Henning, J. E., Kohler, F. W., Robinson, V. L., & Wilson, B. (2009). *Improving teacher quality:*

Using the teacher work sample to make evidence-based decisions. Lanham, MD:

Rowman & Littlefield Education.

International Literacy Association. (2015). *Collaborating for success: The vital role of content*

teachers in developing disciplinary literacy with students in grades 6–12 [Position

statement]. Retrieved from [http://literacyworldwide.org/docs/default-source/where-we-](http://literacyworldwide.org/docs/default-source/where-we-stand/ccss-disciplinary-literacy-statement.pdf?sfvrsn=8)

[stand/ccss-disciplinary-literacy-statement.pdf?sfvrsn=8](http://literacyworldwide.org/docs/default-source/where-we-stand/ccss-disciplinary-literacy-statement.pdf?sfvrsn=8)

- Keller, B. T., & Enders, C. K. (2017). *Blimp Software Manual (Version Beta 6.8)*. Los Angeles, CA.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York, NY: Longman.
- Konstantopoulos, S. (2014). Teacher effects, value-added models, and accountability. *Teachers College Record, 116*(1). Retrieved from <http://www.tcrecord.org/content.asp?contentid=17290>
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis, 25*, 287-298. doi:10.2307/3699496
- Langman, J. (2003). The effect of ESL-trained content-area teachers: Reducing middle-school students to incidental language learners. *Prospect, 18*(1), 14-26.
- Lara-Alecia, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The effect of an instructional intervention on middle school English learner's science and English reading achievement. *Journal of Research in Science Teaching, 49*(8), 987-1011.
- Lavery, M. R. (2012). TWS GraphMaker [Computer software]. Orlando, FL: M. R. Lavery.
- Lavery, M. R. (2017). *Using aggregate unit pre/post assessment results across classrooms: A Monte Carlo study proof of concept in program evaluation*. Paper presented at the Modern Modeling Methods Conference, Storrs, CT.
- League of United Latin American Citizens v. Florida Department of Education. (1990). Retrieved from <http://www.fldoe.org/core/fileparse.php/7582/urlt/Consent-Decree.pdf>
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781119013563

Llosa, L., Lee, O., Jian, F., Haas, A., O’Conner, C., Van Booven, C. D., Kieffer, M. J. (2016).

Impact of large-scale science intervention focused on English language learners.

American Educational Research Journal, 53(2), 395-424. doi:

10.3102/0002831216637348

Long, M. (1996). The role of linguistic environment in second language acquisition. In W.

Ritchie & T. Bhatia (Eds.). *The new handbook of second language acquisition* (pp.

413–468). San Diego, CA: Academic Press.

Lucas, T. (2011). *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators*. New York, NY: Routledge.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for

value-added modeling of teacher effects. *Journal of Educational and Behavioral*

Statistics, 29(1), 67-101. doi:10.2307/3701307

McGraner, K.L., & Saenz, L. (2009). *Preparing teachers of English language learners*.

Washington, D.C.: National Comprehensive Center for Teacher Quality.

Meng, X.-L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed

data sets. *Biometrika*, 79(1), 103-111. doi:10.2307/2337151

National Center for Education Statistics. (2015a). *Nation’s report card: Mathematics 2015*.

Washington D.C.: U.S. Department of Education, Institute of Education Sciences.

Retrieved from https://www.nationsreportcard.gov/reading_math_2015/#?grade=8

National Center for Education Statistics. (2015b). *Nation’s report card: Reading 2015*.

Washington D.C.: U.S. Department of Education, Institute of Education Sciences.

Retrieved from https://www.nationsreportcard.gov/reading_math_2015/#?grade=8

- National Center for Education Statistics. (2015c). *The condition of education 2015, English Language Learners* (NCES 2015-144). Washington D.C.: U.S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/programs/coe/pdf/coe_cgf.pdf
- National Clearinghouse for English Language Acquisition. (2008). *Educating English language learners: Building teacher capacity roundtable report Vol. III. State requirements for pre-service teachers of ELLs*. Available at <http://www.ncela.us/publications>.
- 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.
- Nutta, J.W., Mokhtari, K., Strebel, C., Mihai, F., & Crevecoeur-Bryant, E. (2014). *Educating English learners: What every classroom teacher needs to know*. Cambridge, MA: Harvard Education Press.
- Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59(1), 1-17.
- Rosselli, H., Girod, M., & Brodsky, M. (Eds.). (2011). *Connecting teaching and learning: History, evolution, and case studies of teacher work sample methodology*. Lanham, MD: Rowman & Littlefield.
- Samson, J. F. & Collins, B. A. (2012). *Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness*. Washington, D.C.: Center for American Progress.

- Santau, A. O., Maerten-Rivera, J. L., Huggins, A. C. (2011). Science achievement of English language learners in urban elementary schools: Fourth-grade student achievement results from a professional development intervention. *Science Education, 95*, 771-793.
- Saunders, W. Goldenberg, C. & Marcelletti, D. (2013). English language development: Guidelines for instruction. *American Educator, 37*(2), 13-25.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*(4), 323-355. doi:10.3102/10769986023004323
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage Publications.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 165-179). Rowley, MA: Newbury House.
- Vogt, W. P., Gardner, D. C., & Haeffele, L. M. (2012). *When to use what research design*. New York, NY: Guilford Press.
- Wright, P., White, J. T., Sanders, W. L., & Rivers, J. C. (2010). *SAS® EVAAS® statistical models*. SAS Institute Inc. Cary, NC. Retrieved from <http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf>
- Yen, W. M., & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education, 4*(3), 209-228.

Tables

Table 1

One Plus Model of English Learner Infusion

		Program Outcomes	
		EL Qualified for Academic Subjects (Group 1)	EL Qualified for Language Arts (Group 2)
Program Elements	EL Embedded Courses	Majority of core teacher education and pedagogical content courses embed a focus on English learners	All Language Arts courses deeply and broadly embed a focus on English learners
	EL Specific Courses	One EL specific course on adapting curriculum, instruction, and assessment for ELs including assignments linked to service learning, or early field experiences/ student teaching in certification area	One additional EL specific course on second language development and educational linguistics applied to language arts instruction for ELs including assignments linked to student teaching/internship in language arts certification area
	Field/ Clinical Experiences	Tutoring or practice teaching EL(s) as part of service learning, or early field experiences/ student teaching in certification area	Teaching one or more ELs throughout student teaching/ internship in language arts
	Assessment	Portfolio assignments from EL-embedded courses and the EL-specific course, evaluation of teaching one or more ELs during a field or clinical experience in certification area (TWS)	Additional portfolio assignments from 2nd EL-specific course, evaluation of teaching one or more ELs throughout student teaching/ internship in language arts (TWS)
		Plus +	

Note: Table adapted from Nutta, J.W., Mokhtari, K., & Strelbel, C. (2012). *Preparing every teacher to reach English learners: A practical guide for teacher educators*. Cambridge, MA: Harvard Education Press.

Table 2

Number of Courses (Number of Assignments in Parentheses) with English Learner-Related Coursework for Groups 1 and 2

	Group 1			Group 2		
	Mathematics	Science	Social Studies	English Language Arts	Elementary	Early Childhood
Core Classes	5 (12)	5 (12)	5 (12)	5 (12)	5 (12)	5 (12)
EL Standalone	1 (3)	1 (3)	1 (3)	2 (11)	2 (11)	2 (11)
Specialization	0 (0)	0 (0)	0 (0)	5 (18)	8 (19)	9 (27)
Total	6 (15)	6 (15)	6 (15)	12 (41)	15 (42)	16 (50)

Note: EL standalone courses are devoted entirely to English learner content (second language acquisition, methods of teaching ESOL, English language development) while the other courses listed contain designated materials and related assignments that infuse EL content into the content typically associated with those courses. Teacher candidates in Group 2 are required to work with ELs during internship, while Group 1 teacher candidates are not.

Table 3

Observed Missingness for Student-Level Data

Variable	<i>n</i> (%)
Pre	38 (0.5%)
Post	64 (0.8%)
EL	3 (< 0.1%)
SWD	54 (0.7%)
LowSES	1252 (15.0%)
Minority	75 (0.9%)

Table 4

Mean Number and Percent of Students per Cluster by Demographic Category and Teacher Candidate Group

Subgroup	Group 1		Group 2		All Teacher Candidates	
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
All Students	58.0 (31.2)	—	21.9 (15.2)	—	28.9 (24.0)	—
ELs	6.9 (10.5)	12.3% (13.5%)	3.7 (4.1)	16.7% (12.4%)	4.3 (6.0)	15.9% (12.7%)
SWD	7.6 (6.3)	15.1% (14.3%)	2.9 (3.6)	13.3% (13.6%)	3.8 (4.7)	13.7% (13.8%)
LowSES	25.3 (19.6)	44.4% (25.6%)	11.5 (10.3)	54.5% (29.2%)	14.0 (13.5)	52.7% (28.8%)
Minority	31.4 (19.6)	55.2% (21.1%)	12.5 (10.7)	57.0% (22.1%)	16.2 (14.9)	56.6% (21.9%)

Note: Group 1 consists of $n = 56$ teacher candidates (who taught $n = 3247$ students) in mathematics education, science education, and social science education programs. Group 2 consists of $n = 232$ teacher candidates (who taught $n = 5079$ students) in elementary education and English language arts education programs. Imputations for records with missing data are not included in the descriptive statistics reported here.

Table 5

Solution for Fixed Effects Output for Final Model

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	61.6463	1.7676	8281.4	34.88	<.0001
Pre	0.3451	0.0091	8003.1	38.10	<.0001
EL	-1.5743	0.4849	7690.7	-3.25	0.0012
SWD	-4.7107	0.4950	6401.1	-9.52	<.0001
LowSES	-2.4494	0.4334	355.7	-5.65	<.0001
Minority	-1.1316	0.3664	6187.8	-3.09	0.0020
Group	3.1826	1.9255	8290.2	1.65	0.0984

Note: Estimates shown were pooled across $m = 25$ imputations of missing data for $n = 8326$

K-12 students taught by $n = 288$ teacher candidates.

Table 6

Solution for Fixed Effects of Post-Assessment Normal Curve Equivalence (NCE) Score Analysis

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	32.7450	0.6878	7698.3	47.61	<.0001
Pre	0.3966	0.0101	8152.5	39.41	<.0001
EL	-2.1008	0.6118	8009.3	-3.43	0.0006
SWD	-5.6623	0.6189	6610.4	-9.15	<.0001
LowSES	-2.1227	0.4836	624.3	-4.39	<.0001
Minority	-1.2088	0.4578	6473.0	-2.64	0.0083
Group	0.3504	0.4232	7846.8	0.83	0.4077

Note: Estimates shown were pooled across $m = 25$ imputations of missing data for $n = 8326$

K-12 students taught by $n = 288$ teacher candidates.

Table 7

Solution for Fixed Effects of Pre-Assessment Normal Curve Equivalence (NCE) Score Analysis

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	53.8725	0.4650	5571.0	115.86	<.0001
EL	-5.3163	0.6653	8118.5	-7.99	<.0001
SWD	-5.8387	0.6693	7668.1	-8.72	<.0001
LowSES	-3.3394	0.5193	800.8	-6.43	<.0001
Minority	-1.8652	0.4984	7045.3	-3.74	0.0002
Group	0.6470	0.4636	7468.4	1.40	0.1628

Note: Estimates shown were pooled across $m = 25$ imputations of missing data for $n = 8326$

K-12 students taught by $n = 288$ teacher candidates.

Table 8

Mean Parameter Estimate, Effect Size, and Rate of Significance for Dichotomous Predictors

Using Randomly Selected Sub-Samples of Data

Effect	Final Model			NCE Model		
	$\bar{\beta}$	\bar{d}	Sig.	$\bar{\beta}$	\bar{d}	Sig.
EL	-1.11 (0.89)	-0.16 (.12)	7.9%	-1.34 (1.07)	-0.19 (.15)	8.4%
SWD	-4.10 (0.81)	-0.58 (.11)	99.5%	-4.49 (0.95)	-0.64 (.13)	97.4%
LowSES	-2.81 (0.62)	-0.55 (.13)	91.9%	-2.47 (0.58)	-0.48 (.12)	68.6%
Minority	-1.22 (0.51)	-0.27 (.11)	18.4%	-1.15 (0.58)	-0.26 (.13)	7.1%
Group	3.05 (2.69)	0.57 (.51)	11.5%	0.35 (0.28)	0.07 (.05)	0.0%

Note: Results shown are the means (with standard deviations in parentheses) of analyses comparing the pre and post assessment results for the students of 30 randomly selected teacher candidates from each group across 1000 runs. Sig. displays the percent of runs for which the parameter was found significant at $\alpha = .019$ ($\alpha = .051$ for *Group*).

Figures

Microteach 1 Lesson Plan – Direct Teach Template
(Revised January 2014)

Name:	Grade Level:
Subject:	Unit Title:
Lesson Title:	Estimated Time:

Standard/s (Sunshine State, Next Generation Sunshine State Standard/s (NGSSS) or Common Core State Standard/s (CCSS)): [2 pts.]

Learning Goal/Target/Objective – Based on standards: [3 pts.] (What should students know and be able to do as a result of this lesson?)

Essential/Guiding Question(s): [3 pts.] (Higher-order thinking level in student-friendly language)

Content: [3 pts.] (Outline what you are going to teach)

Key Vocabulary: [2 pts.]

Procedures: [7 pts.] Insert ESOL accommodations for Edith, Edgar and Tasir in **BOLD, CAPITAL LETTERS** [3 pts.]

1. **Activate prior knowledge/Anticipatory Set/Hook:** What prior knowledge should students have for this lesson? How will you gain their attention?
2. **Input:** Teaching Phase: How will you present the concept or skill to your students?
 - a. Key vocabulary instruction
 - b. Teaching of the concept
3. **Modeling:** Showing, explaining, demonstrating to the students how to do something that they will then be expected to do.
4. **Check for understanding:** Whatever you do to determine students' comprehension of what has been taught thus far (Q & A, cognitive processing strategies, graphic organizers, use of scales, etc.)
5. **Guided practice:** What activities or exercises will the students complete with teacher and/or peer guidance?
6. **Independent practice:** What activities or exercises will students complete to reinforce the concepts and skills developed in the lesson?
7. **Closure:** How will you assist your students in organizing the knowledge gained in the lesson? (Q & A, exit slip, game, cognitive processing strategies, etc.)

Assessment and Monitoring: [3 pts.] How will you assess the students' attainment of the lesson objectives? The assessment must be stated in measurable terms (condition, performance, criteria).

- Formative Assessment and/or Summative Assessment:

Materials/Resources: [2 pts.] All materials and equipment needed to teach this lesson.

References: [1 pt.] Any sources (e.g., textbooks, Internet sites, etc.) used in planning this lesson.

Adaptations for Unique Student Needs: [6 pts.] (EL, Special Education, Gifted, Students who lack support for school)
Explain EL accommodations made for Edith, Edgar and Tasir. Explain why the accommodation was made; how it will help the EL student.

Figure 1. Example of an infused ESOL assignment used in the One Plus model.

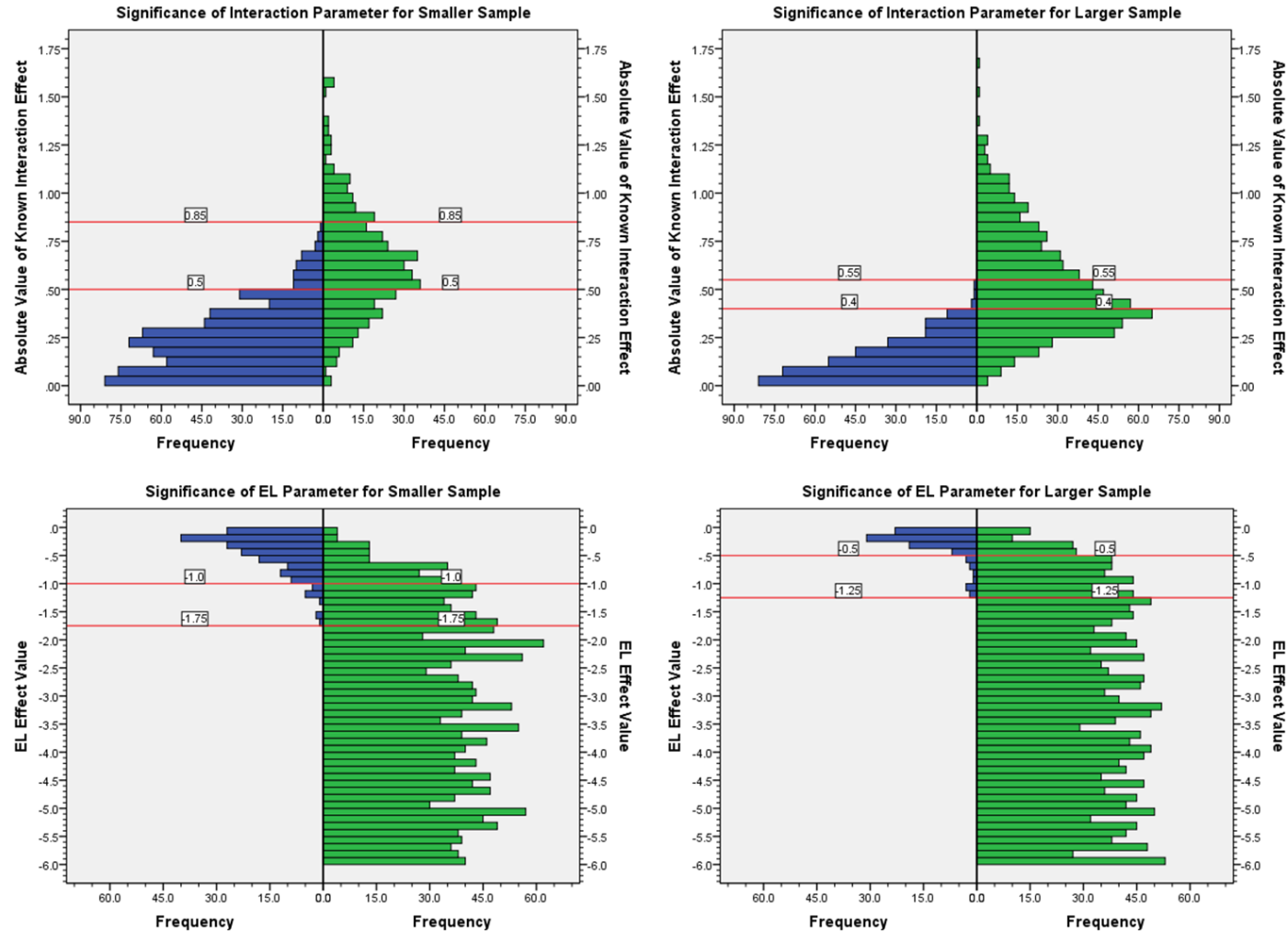


Figure 2. Significance of EL Parameter and $EL*Group$ Interaction Parameter in Simulation Study. The left panels indicate simulation runs in which the parameter was non-significant, while the right panels indicate parameter significance. The first horizontal reference line indicates the magnitude of effect at which the error rate becomes acceptable, while the second horizontal reference line indicates the point at which the error rate falls to zero.