PROVIDING EQUAL ACCESS TO ENGLISH LEARNERS IN EDUCATIONAL SETTINGS

by

Tuba Gezer

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Educational Research, Measurement and Evaluation

Charlotte

2021

Approved by:

_____
Dr. Claudia Flowers

_____
Dr. Richard Lambert

_____
Dr. Stella Kim

_____
Dr. Valerie Mazzotti

ABSTRACT

TUBA GEZER. PROVIDING EQUAL ACCESS TO ENGLISH LEARNERS IN EDUCATIONAL SETTINGS (UNDER THE DIRECTION of Dr. CLAUDIA FLOWERS)

This three-article format dissertation examined fairness in testing English Learners (EL) in K-12 schools in the United States (US). In the first article, a meta-analysis was conducted to summarize EL computer-based testing (CBT) accommodations research on the validity and effectiveness of accommodations. Eight studies out of 292 studies met the inclusion criteria. The results indicated that CBT accommodations did not influence non-EL test scores. There was a small and statistically significant (.12 SD) improvement in EL test scores for EL students with CBT accommodations. The second article focused on predictors of postschool outcomes (PSO) for EL students with disabilities. Data from NLTS-2 was used to examine the direct and indirect effects of predictor factors, which included adaptive behavior, parent expectations, and transition services, on PSO. The results suggested that transition planning totally mediated the effects of adaptive behaviors and parent expectations on PSO. The third article used EL large-scale assessment data to investigate measurement invariance (MI) and prediction invariance (PI) across ethnicity and years in the US schools. A multigroup structural equation model was used to simultaneously examine MI and PI. The results suggested MI and PI were retained between Hispanic and non-Hispanic EL students in 3rd and 8th grades, but there was a partial weak MI between Hispanic and non-Hispanic in 10th grade. Although there was MI and PI for 3rd grade across the years in the US schools, only partial measurement invariance was attained for 8th and 10th graders. These results suggest test scores may not be comparable for EL students in the US three years or less and those EL students in the US more than three years among 8^{th} and 10^{th} grade students. Implications and recommendations for future research are discussed.

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my patient and supportive academic adviser and dissertation chair, Dr. Claudia Flowers. I am grateful every day that you were the one who guided me through this process. I honestly could not ask for a better advisor and dissertation chair. Thank you for making time for me. Your kindness and encouragement have made the dissertation process so much easier than I could have ever imagined, even though the majority of this work was completed during a global pandemic. I admire you for your hard work, kindness, knowledge, life-work balance. I hope to be like you when I became a faculty. I must say that you set the bar so high, though, and I promise to treat my students the way you have treated me during my dissertation process.

I would like to express my deepest gratitude to the committee members of my dissertation, Drs. Richard Lambert, Stella Kim, and Valerie Mazzotti for their constructive feedback. I appreciate the time, patience, and intellectual contribution that you added to improving my dissertation. Thank you for helping me to complete my dissertation successfully.

I sincerely thank Educational Leadership faculty Drs. Sandra Dika, Jae Hoon Lim, Chuang Wang, Carl Westine, Xiaoxia Newton, Lynn Ahlgrim-Delzell, Florence Martin, Ayesha Sadaf, Beth Oyarzun, and Ryan Miller for their instruction, guidance. I appreciate everything I learned from you regarding content knowledge, work ethic, research experiences, and publication process. I would like to thank my dear ERME friends/colleagues, Ting, Dave, Tong, Liz, Kristin, Yi, Nick for their friendship, assistance. I treasure our time together in classrooms, COED 220, ELSO events, or on Zoom. I thank Dr. Brian Gong for helping to find my research interest and Dr. Roslyn Mickelson for teaching me the background of EL education in the US. I thank Dr. Aziz Harman and Ali Karakas for their support in starting my education in the US.

I would like to express my heartfelt appreciation to my family, especially to my mother, Fatma Gezer, my father, Hamdullah Gezer, for their unconditional love and prayers. I sincerely thank my siblings Emine, Said, Saliha, Semsettin, Mehmet, Abdullah, Zehra, and Muhammed Nurullah for their love, encouragement, support, and time with me on the phone. Also, thank you all for taking care of my parents and one another while I was studying in the United States. I would like to honor the memory of my grandmother, Saya Sultan Gezer, who prayed for me to be a doctor. Although she meant medical doctor, I like to think that Allah accepted her prayers in a slightly different form. I thank my grandparents, Semsettin Gezer, who I have never met, and Bedri Kaya, who is in my warm childhood memories. Although both are no longer with us, I know I got the courage to move to the other side world from them. I appreciate their audacity, and their memories will always live with me. I would like to thank my nieces and nephews, Sena Nur, Sema Nur, Muhammed, Beyza, Semih, Elif, Bilal, Hira, Alp, Mehmet, Ali, Dara, and Umut, who have been a constant source of joy, support, and happiness.

I thank my friends who supported me in this process. I appreciate our phone calls, facetime, long conversations, laughs, meals, trips, complaints, study sessions, encouragement, and supports. You all made my life better; thank you for sticking around regardless of the distance between us. I specifically want to thank Elife and her family. I would not know you will become such an incredible friend when we got on a plane together seven years ago without really knowing one another. Thank you for checking on me, especially for your hourly messages to check my progress at the last stage of my dissertation writing process.

I appreciate the YLSY program by the Turkish Republic, the Ministry of Education for supporting students to receive higher education abroad. Once again, I thank my family, friends, and mentors. My love for you kept me going, and your love kept me sane in this process.

DEDICATION

This dissertation is dedicated to all the people in my life who touch my heart. I am who I am because of your unconditional love and constant support. I am truly thankful for having you in my life. I, also, dedicate this research to all language minorities around the world.

Bele, em dikarin!

Evet, yapabiliriz!

Si, se puede!

Yes, we can!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ESEA	The Elementary and Secondary Education Act
NCLB	The No Child Left Behind Act
ELs	English Learners
ESSA	The Every Student Succeeds Act
UD	Universal Design
CBT	Computer-based Testing
NAEP	National Assessment of Educational Progress
SWD	Student with Disabilities
IDEA	The Individuals with Disabilities Education Act
NLTS 2	the National Longitudinal Transition Study 2
PSO	Postschool Outcomes
CI	Confidence interval
SEM	Structural equation modeling
RMSEA	Root mean square error of approximation
CFI	Comparative fit index
SRMR	Standardized root mean squared residual
MI	Measurement invariance
PI	Prediction invariance
ELA	English language art
ELP	English language proficiency
WIDA	World-Class Instructional Design and Assessment
MGSEM	Multigroup structural equation modeling
LTEL	Long-term English learners

INTRODUCTION

Approximately 62 million people in the United States speak a language other than English at home, and 41% of them are not proficient in English (Zong & Batalova, 2015). From 1990 to 2013, the proportion of individuals who were limited English proficient increased by 80%. The analysis of the labor force indicated that nearly 10% of the workforce is limited English proficient, and the English-proficient workforce earns 25-40% more than limited English proficient working adults (Wilson, 2014). In the United States, English proficiency is an essential skill for immigrants and native-born people.

According to the Elementary and Secondary Act of 1965, as amended by Every Student Succeeds Act (ESSA) in 2015, the term English Learner (EL) was defined as someone who is between 3-21 years old, enrolled in elementary or secondary school, and who was not born in the United States or had a native language other than English. As the EL population increases in schools, meeting fairness testing standards, and providing equal educational opportunities to EL gain critical importance. By 2025, the proportion of EL is expected to reach 25% (National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs, 2007). A recent study by McFarland and his colleagues (2017) reported that there were 4.6 million (9.4%) ELs attending school in 2014-2015. In some states such as California, this proportion could go up to 22%. The same study showed the ratios of EL in urban cities (17%) and in suburban areas (9%). However, the analysis of demographic changes in rural areas showed that rural EL enrollment increased by nearly 72% from the 1999-2000 school year to the 2014-2015 academic year (Johnson et al., 2018). Considering the EL population increases, educational resources and programs are needed to ensure EL students can successfully meet the high expectations needed to transition into jobs and postsecondary education.

The Civil Rights Act of 1964, subsection 601 prohibits any discrimination based on "race, color, or national origin" in federal financial assisted programs/activities and the implementations of the Department of Health, Education, and Welfare. Regarding the case of *Lau v. Nichols (1974)*, Supreme Court decided that San Francisco Unified School District violated the *Civil Rights Act* because the school district did not provide meaningful opportunities to Chinese speaking minority students to be a part of public education programs (p.567). The Supreme Court required the school district to offer educational opportunities for language minority students, such as bilingual education. Despite the Supreme Court decisions, many educators lack the understanding needed to implement high-quality programs well enough to help all students. The case of *Lau v. Nichols* was the beginning of EL students' civil rights, and research-based practices are needed to provide all EL access to high-quality education.

EL Population Demographic Characteristics

The EL population has diverse demographic characteristics such as race and native language, disabilities, socioeconomic status, and immigration status. Among EL students, Spanish (77%) is the most common home language while a small proportion of EL speak Arabic (2%) and Chinese (2%; McFarland et al., 2017). Also, the researchers reported that 14% of EL students are identified as EL with disabilities. The majority of EL with disabilities (57%) is identified as having learning disabilities (Trainor et al., 2019). Besides, the EL population is more likely to be in low socioeconomic status as 14% of EL experience homelessness, and 47% of them attend high poverty schools in the 2014-2015 academic year (McFarland et al., 2017). Although the majority of EL is born in the United States, some of them immigrated to the US (Flores et al., 2017). Thus, many EL students' academic progress could have been interrupted due to mobility (Calderón et al., 2011).

Another unique aspect of the EL population is that the characteristics of this population are changing. EL students who reach the language proficiency exit EL status while newcomers join EL populations with a variety of English background. According to ESSA (2015), newcomers are defined as someone who comes to the United States in the last 12 months, and these students can be excluded from one reading or English Language Art assessment. Reflecting on the diversity and the mobility of EL students, determining the educational needs of EL students could be challenging yet imperative to examine. This study, thus, investigates fairness and equal educational access for EL students.

Educational Policies regarding EL Population

With the execution of the No Child Left Behind (NCLB) Act of 2001, states have to test students in Reading and Math from 3rd grade to 8th grade and report test results publicly, including the disadvantaged student groups as Black, special education, and EL. Therefore, EL English language proficiency and academic achievement gained more attention. A consequence of NCLB was that many schools failed to reach the targeted goal of 100% proficiency. ESSA, the current federal law, gave more flexibility to the states regarding reaching the desired achievement level compared to NCLB. ESSA still requires school systems to set high expectations for all students and report results by student subgroups, including EL students. The following section explains fairness standards and equal educational access for the EL population to meet the high expectations.

Fairness in Testing: Accessibility to Constructs

From a testing perspective, fairness is defined as providing all test takers with an equal opportunity to demonstrate what they know and can do on the construct of interest (American Educational Research Association [AERA], American Psychological Association [APA], &

National Council on Measurement in Education [NCME], 2014). Fairness "is a central issue in achieving valid test results" (Plake & Wise, 2014, p.7).

States are required to make valid and reliable inferences from assessments and provide evidence of validity and reliability. Young et al. (2008) articulated that state content assessments have to be fair and valid so that test results should be based on content knowledge, not irrelevant constructs like language proficiency. Accurate assessment of EL students is an essential aspect of monitoring their performance (Pennock-Roman & Rivera, 2011). EL students, therefore, may receive test accommodations until reaching English language proficiency. With the increasing use of technology in educational settings, computer-based testing (CBT) becomes a common practice, especially after the Smarter Balanced Assessment Consortium and the Partnership for Assessment of Readiness for College and Careers (PARCC) adopted CBT for national assessment and accountability systems. Since the 2015-2016 academic school year, the majority of students in K-8 has taken the online summative assessment; in fact, only 15 % of the assessment was offered only on a paper-and-pencil format (Strategies, 2015). Adaptation of CBT may allow the use of multiple accommodations simultaneously, such as combining pop-up glossary and read-aloud accommodations (Russell et al., 2009). The first study, in chapter II, conducted a meta-analysis to investigate the effectiveness of computer-based test accommodations for EL students. Meta-analysis may minimize the sampling error, random error of individual studies, and summarize the research to reveal the collective conclusion (Borenstein et al., 2009).

Equal Educational Opportunities

The United States' Constitution requires that all students be given equal educational opportunities no matter their race or ethnic background. The Supreme Court declared in Brown

v. Board of Education that education "is a right which must be made available to all on equal terms." Education is a highly valued good for both the individual and society. As the number of EL students is increasing, the number of EL with disabilities is growing concurrently. Maintaining equal educational opportunities, therefore, has critical value. The research suggested that schools need to concentrate on language and special education services equally in order to provide adequate educational opportunities to EL with disabilities (Kangas, 2017). However, some school districts could enforce a formal/informal de facto policy, which does not allow EL with disabilities to receive language and special education services even though this is against the federal law (US Department of Justice & US Department of Education, 2015).

There is a significant difference between EL with disabilities and students with disabilities in terms of employment outside of the home (Trainor et al., 2019). The researchers indicated EL with disabilities, and their parents have lower postsecondary education expectations compared to students with disabilities (SWD) students and their parents. Both student groups' expectations are significantly lower than general education students and their parents' expectations. While 41% of general education students expected to attain a 4-year college, only 31% of SWD and 25% of EL with disabilities are expected to attain a 4-year college. Considering disparities between EL with disabilities and SWD's post-school transition, the aim of second study is to analyze the effective predictors of post-school transition for EL with disabilities.

Fairness: Predictive and Measurement Invariance

Prediction invariance (PI) maintains the uniform interpretation of test results regardless of demographic characteristics such as gender, race, and language status (AERA, APA & NCME, 2014). Examination of PI research suggested that observed scores of test results, SAT and high

school GPA may lead to overprediction of first-year college performance for ethnic groups (Aguinis et al., 2016; Berry & Zhao, 2015; Mattern & Patterson, 2013) and underprediction of females' college grades compare to males (Fischer et al., 2013; Keiser et al., 2016). Considering a typical aim of standardized tests is to predict academic or job performance (Cleary, 1968; Humphreys, 1952), PI becomes crucial because overprediction leads to unfair benefits, and underprediction causes unfair penalties (Culpepper et al., 2019).

Measurement invariance (MI) is a statistical property of measurement that indicates the same underlying construct is being measured across multiple groups and times (Putnick & Bornstein, 2016). The researchers further emphasized that MI has a critical value to psychological research because it is a requirement of comparing group means. Millsap (2007) stated that when tests are used for selection, prediction gains more attention, which inadvertently leads to overlooking measurement concerns in the testing process. The author further suggested analyzing measurement invariance and prediction invariance using the same data.

Jonson et al. (2019) reviewed 18 academic and intelligence tests to investigate to what extent fairness standards (AERA, APA & NCME, 2014) were practiced. The authors found that preliminary sampling likely to examine fairness standards for age, gender, and race/ethnicity subgroups. While the SWD subgroup is usually included in the standardization sample, EL students have been overlooked in both preliminary and standardization samplings. The author concluded that since SWD and EL students are not represented in preliminary sampling, the further fairness issues of reliability and validity may occur. Since MI and PI are necessary aspects of test fairness, Chapter IV leans towards the application of multi-group structural equation modeling to assess measurement and predictor invariance simultaneously for EL high-stake testing data.

Research Questions

The purpose of this study is to examine the test fairness and equal educational access for EL in the United States. The following research questions have guided this study.

1. Do EL test accommodations on CBT influence non-EL students' academic performance? To what extent are CBT test accommodations effective in improving EL academic performance? What factors influence the effectiveness of EL test accommodations?
2. To what degree do transition planning, family factors and adaptive behaviors predict postschool outcomes for EL with disabilities?
3. Are English language proficiency measures (i.e., listening, speaking, reading, and writing) invariant across ethnicity and time in the US schools? Are English language proficiency measures of prediction of ELA achievement invariant across ethnicity and time in the US schools? Is there a difference between a latent score and observed score prediction of ELA achievement?

Dissertation Overview: The Three Articles

I, Tuba Gezer, is the first author in all three articles in this dissertation. My dissertation chairs and a committee member are my co-authors in each paper. Therefore, "we" is referred to my co-authors and myself throughout the articles.

Chapter 1 [Article 1]: Effectiveness of English Learners Computer-Based Testing Accommodations: A Meta-Analysis

With the increased number of English learners (ELs) participating in large-scale state testing, there has been an increased focus on fairness in testing for all students. Test accommodations have shown promise in eliminating barriers and providing equal access for all

test takers, and computer-based testing (CBT) allows individual customization of tests with built in accommodations. Most research on testing accommodations focuses on paper-based tests but CBT is the most predominant mode of delivering large-scale state assessments. The purpose of this study is to synthesize research on the validity and effectiveness of CBT accommodations for EL students. Meta-analysis methodology was used to summarize the findings from previous CBT studies. Eight studies out of 292 studies met the inclusion criteria. The results indicated that CBT accommodations did not influence non-EL test scores, suggesting that the construct being measured was not changed because of the accommodation. There was a .12 standard deviation improvement in EL test scores for EL students who had CBT accommodations. The grade level of EL moderated the effectiveness of the accommodation, with elementary students demonstrating higher effects than middle and high school students. The findings of this study are similar to those studies that examined paper-based accommodations. Limitations of the study and future research are discussed.

Chapter 2 [Article 2]: Predicting Postschool Outcomes for English Learners with Disabilities: Secondary Analysis of Data from the National Longitudinal Transition Study-2

Employment, postsecondary education and independent living are the common indicators of a successful postschool transition. There is a large amount of research examining students with disabilities postschool outcomes, but limited research has investigated the unique challenges of English learners (EL) with disabilities. This study used structural equation modeling to examine the relationships between EL with disabilities' postschool outcomes (PSO) and adaptive behaviors, parent expectations, and transition planning. Data from National Longitudinal Transition Study-2 was used to examine the direct and indirect effects of predictor factors on PSO. The results suggested that transition planning mediated the effects between PSO and

adaptive behaviors and parent expectations. This study expanded the literature regarding EL with disabilities' PSO and the importance of parent expectations and transition planning.

Chapter 3 [Article 3]: Applying Latent Variable Approach for Examining Measurement and Prediction Invariance in English Learners Large-scale Assessments

Measurement invariance investigates the consistency of measures for different conditions such as time, population, or method, and prediction invariance maintains the uniform interpretation of test results regardless of demographic characteristics such as gender, race, language status. While the EL population has diverse characteristics among ethnicity, home language, and immigration status, the diversity of the EL population has been overlooked by the accountability systems. Therefore, the purpose of this study is to analyze measurement and prediction invariance simultaneously based on latent scores and compare the results with observed scores prediction using EL large-scale testing data for 3rd, 8th, and 10th graders. The latent score analysis suggested measurement and prediction invariance among Hispanic and non-Hispanic EL students in 3rd and 8th grades, but there was partial weak measurement invariance among Hispanic and non-Hispanic in 10th grade. Although there was measurement and prediction invariance for 3rd grade across the year in the US schools, only partial measurement invariance was attained for 8th and 10th graders. Besides, there was not prediction invariance across the year in the US in 10th grade. The results suggested that observed scores may not be accurately compared across students who have been in the US schools less than three years and students who have been in the US schools for more than three years.

Limitations

There is a number of limiting factors for these collective studies. First, the quality of meta-analysis is depended on the quality of the included studies. Regarding the second study, EL

students are over-represented in learning disability categories, which raise the accuracy of identifying EL with disabilities. As we analyzed the secondary data, NLTS 2, we assumed EL students have identified as student with disabilities accurately. While the third study analyzed measurement and predictive invariance using real data, the results cannot be generalized to all EL in the United States because the data was obtained from one state.

Significance of the Study

Growing EL population increases the importance of fairness in testing by accessibility to construct and eliminating measurement error along with offering equal educational opportunities to EL students. This study highlighted the importance of fairness in educational setting from accessibility, equal educational opportunities and measurement and prediction invariance models. The results of this study benefit teachers, parents, department of education staff, and EL students to expand the fairness in testing and equal educational access.

Aside from benefiting the EL population in the US, with growing immigration and globalism, EL education becomes a significant part of international education. Liasidou (2013) suggested that English speaking western countries need to accommodate diverse learner population as the number of immigrant students is increasing. For instance, the growing EL population in the United Kingdom and the Kurdish and Turkish student population in Germany demonstrate the need to accommodate language minorities as an important aspect of international education (Abedi, 2013).

References

Abedi, J. (2013). Validity issues in designing accommodations for English language learners. *The Routledge handbook of language testing* (pp. 62-76).

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2016). Differential prediction generalization in college admissions testing. *Journal of Educational Psychology*, 108, 1045–1059.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Berry, C. M., & Zhao, P. (2015). Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance. *Journal of Applied Psychology*, 100, 162–179.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK: John Wiley & Sons.

Calderón, M., Slavin, R. E., & Sánchez, M. (2011). Effective instruction for English learners. *Future of Children, 21*, 103–127.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, 5, 115–124.

Culpepper, S. A., Aguinis, H., Kern, J. L., & Millsap, R. (2019). High-stakes testing case study: A latent variable approach for assessing measurement and prediction invariance. *Psychometrika*, 84(1), 285-309.

Every Student Succeeds Act. (2015). *Every Student Succeeds Act of 2015*, Pub. L. No 114 – 95 (2015 - 2016).

Fischer, F. T., Schult, J., & Hell, B. (2013a). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational Psychology*, 105, 478–488.

Flores, S. M., Park, T. J., Viano, S. L., & Coca, V. M. (2017). State policy and the educational outcomes of English learner and immigrant students: Three administrative data stories. *American Behavioral Scientist, 61*(14), 1824-1844.

Humphreys, L. G. (1952). Individual differences. *Annual Review of Psychology*, 3, 131–150.

Johnson, J., Ohlson, M. A., & Shope, S. (2018). Demographic changes in rural America and the implications for special education programming: A descriptive and comparative analysis. *Rural Special Education Quarterly*, *37*(3), 140-149.

Jonson, J. L., Trantham, P., & Usher‐Tate, B. J. (2019). An Evaluative Framework for Reviewing Fairness Standards and Practices in Educational Tests. *Educational Measurement: Issues and Practice,* 38(3), 6-19.

Kangas, S. E. N. (2017). "That's where the rubber meets the road": The intersection of special education and dual language education. *Teachers College Record, 119*(7).

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, 101, 569–581.

Liasidou, A. (2013). Bilingual and Special Educational Needs in Inclusive Classrooms: Some Critical and Pedagogical Considerations. *Support for Learning, 28*(1), 11–16.

Lipscomb, S., Haimson, J., Liu, A.Y., Burghardt, J., Johnson, D.R., & Thurlow, M.L. (2017). *Preparing for life after high school: The characteristics and experiences of youth in special education. Findings from the National Longitudinal Transition Study 2012. Volume 1: Comparisons with other youth: Full report (NCEE 2017-4016)*. Washington,

DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Mattern, K. D., & Patterson, B. F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, and Pierce (2010). *Journal of Applied Psychology*, 98, 134–147.

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., and Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017- 144). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144.

Millsap, R. E. (2007). Invariance in measurement and prediction revisited. *Psychometrika,* 72(4), 461-473.

National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs. (2007). Retrieved from http://sbo.nn.k12.va.us/esl/documents/ ncela_fast_faqs.pdf

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Harvard Education Press.

No Child Left Behind Act of 2001, PL 107-110, 20 USC. (2002).

Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice,* 30(3), 10-28.

Plake, B. S., & Wise, L. L. (2014). What is the role and importance of the revised AERA, APA, NCME Standards for Educational and Psychological Testing*? Educational Measurement: Issues and Practice, 33*(4), 4-12.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review, 41*, 71-90. http://dx.doi.org/10.1016/j.dr.2016.06.004

Russell, M., Hoffmann, T., & Higgins, J. (2009). Meeting the needs of all students: A universal design approach to computer-based testing. *Innovate: Journal of Online Education*, 5(4).

Strategies, E. (2015). *Pencils down: The shift to online & computer-based testing*.

Trainor, A. A., Newman, L., Garcia, E., Woodley, H. H., Traxler, R. E., & Deschene, D. N. (2019). Postsecondary Education-Focused Transition Planning Experiences of English Learners with Disabilities. *Career Development and Transition for Exceptional Individuals, 42*(1), 43–55.

US Department of Justice & US Department of Education. (2015, January). *English learner (EL) dear colleague*. Retrieved from https://www2.ed.gov/about/offices/list/ocr/letters/colleague-el-201501.pdf

Wilson, J. H. (2014). Investing in English skills: The limited English proficient workforce in US metropolitan areas. *The Brookings Institution.*

Young, J. W., & King, T. C. (2008). *Testing accommodations for English language learners: A review of state and district policies.* Research Report No. 2008-6. ETS RR-08-48. College Board.

Zong, J., & Batalova, J. (2015). The limited English proficient population in the United States. *Migration Information Source*, e126.

CHAPTER 1 [ARTICLE I] EFFECTIVENESS OF ENGLISH LEARNERS COMPUTER-BASED TESTING ACCOMMODATIONS: A META-ANALYSIS

Gezer, T., Flowers, C. & Lambert, R. (Revised & Resubmitted). Effectiveness of English learners computer-based testing accommodations: A meta-analysis. *Educational Assessment*

In 2001, the federal government reauthorized the Elementary and Secondary Education Act (ESEA), which increased state accountability for students with disabilities and students who are English learners (ELs). More recently, Every Student Succeeds Act (ESSA) required states to create goals and interim measures for English language proficiency (United States Department of Education, 2015). With approximately 10% of all school children being ELs (McFarland et al., 2018) and with the increased number of ELs participating in state testing systems, there has been an increased focus on fairness in testing students who are ELs.

The Standards for Educational and Psychological Testing emphasizes fairness as a fundamental validity issue that should be addressed in all phases of the testing process (AERA, APA, & NCME, 2014). Designing fair tests requires developers to consider reducing barriers for all examinees during test development, administration, scoring, interpretation, and uses (Thurlow et al., 2009). An important concept in fairness is accessibility. Accessibility focuses on eliminating barriers and providing equal access for all examinees, which allows for score interpretations to have comparable meaning for individuals or groups in the intended population of test-takers (Stone & Cook, 2018). Methods for increasing accessibility and fairness in testing include careful design of assessments and the use of accommodations (Thurlow et al., 2009). Test developers consider the needs of the intended test population in the earliest design stage using universal design (UD) principles. UD principles provide a model for designing accessible tests by focusing on all test takers' needs (AERA, APA, & NCME, 2014), and the use of

accommodations is considered during the assessment development phase. In this study, we are using accessibility and accommodation as defined in the Standards for Educational and Psychological Testing. Accessibility is the notion of providing unobstructed opportunity to all students to demonstrate their ability on the measured construct (AERA, APA, & NCME, 2014, p. 49) and accommodations refers to changes in the test format, test administration, or response procedure while maintaining the original construct (AERA, APA, & NCME, 2014, p.58).

In addition to these test design processes, computer-based testing (CBT) was seen as the test administration mode that can increase access and flexibility to most students' assessments. CBT has made it easier and less expensive to accommodate specific test takers' needs for presenting test items during an assessment by reducing the number of paper-based test forms required to meet specific student's needs. In the U.S. Department of Education's major initiatives (e.g., Race to the Top Assessment Program), the development of CBT was encouraged because of the many positive merits, including built-in accommodations (Thurlow, Lazarus, Albus, & Hodgson, 2010). Unlike paper-based tests, CBT allows individual customization of tests. For example, CBT allows the presentation of an item to vary by font size, magnification, color contrast based on individual student's needs and preferences. Given the ability to customize the presentation and response modes in the CBT environment, test developers can build accommodations into the test design to eliminate the construct irrelevant variance.

While the number of EL students is increasing in the US educational system, ELs' academic performance continues to lag behind their native English-speaking peers. Based on 2017 data from the National Assessment of Educational Progress (NAEP), approximately half of all states reported a decrease in the percentage of grade 4 proficiency rates in mathematics and similar results are noted in reading proficiency rates (NAEP, 2017). The academic under-

performing of EL students is due, in part, to the challenges for ELs in communicating in the English language, especially on core subject tasks (Abedi & Levine, 2013). Test accommodations are one support that has been shown to be effective at narrowing the achievement gap. Research suggests that ELs who received testing accommodations outperform ELs who do not receive accommodation or do not receive appropriate accommodations (Abedi, 2009; Kopriva et al., 2007).

Research has examined EL test accommodations' effectiveness and validity (Abedi, 2009; Abedi et al., 2020; Albus et al., 2005; Johnson & Monroe, 2004; Kopriva et al., 2007). Several meta-analyses summarized the fairness and effectiveness of EL test accommodations (Li & Suen, 2012a; 2012b; Pennock-Roman & Rivera, 2011; Rios et al., 2020) and EL test accommodations in large-scale testing settings (Kieffer et al., 2009). However, little is known about CBT accommodations for ELs even though CBT is the predominant mode of delivering large-scale summative tests in the U.S. educational system. Therefore, the purpose of this meta-analysis is to aggregate data from CBT accommodation studies to evaluate the validity and effectiveness of accommodations for ELs.

Literature Review

Test Accommodations

EL test accommodations are designed to reduce the construct irrelevance variance while increasing the content accessibility for ELs. Some of the common test accommodations on paper-based tests are extended time, dictionaries (e.g., glossary, English dictionary, bilingual dictionary), and linguistically simplified test items (Rivera, 2003; Shafer Willner et al., 2008). Table 1-1 presents the description of EL test accommodations.

Table 1-1

The Description of EL Test Accommodations

Accommodations	Descriptions
Dictionaries	Dictionary accommodations provides the defining the words, but it does not include content-related terms. Dictionary accommodation on CBT provides a simple definition of a word when the mouse is on a word. They can provide visual support or the translation of the word.
Linguistic Modification	Simplifying the language complexity of the test.
Translation	Administering the test in students' native language. It is offered if students are proficient in their native language. The Spanish version of the tests is more common.

Abedi (2006a) examined the difficulties of measuring EL students' content knowledge due to the linguistic complexity in many academic content area tests. He found statistically significant difference between EL and non-EL student groups' measurement error resulting in disadvantages for EL students. Although a large amount of research highlights the achievement gap between EL and non-EL students (Abedi, 2006a; Miley & Farmer, 2017; Polat et al., 2016; Solano-Flores & Trumbull, 2003; Wolf et al., 2008), the source of this achievement gap could be language proficiency, not the content knowledge (Abedi, 2006b). An examination of standardized test scores demonstrated that ELs and native speakers' achievement gap was higher in reading and writing in high language demand items than in lower language demands in mathematics tests (Abedi, Leon, & Mirocha, 2003). Since most content assessments are not designed to measure student linguistic abilities, the language demand on content assessment is a threat to construct validity. Therefore, providing effective and valid test accommodations is critical to reveal EL students' actual performance on content assessments.

Test accommodations should be based on student needs (Francis et al., 2006; Kopriva et al., 2007). For EL students, assigning accommodations is a systematic process to meet students'

linguistic needs. Abedi (2013) suggested that the following conditions for offering appropriate test accommodations to EL students:

(a) Accommodations should be effective in increasing the accessibility of assessment.

(b) Accommodations need to be valid by reducing construct-irrelevant variance.

(c) The benefits of accommodations are dependent on students' backgrounds.

(d) Accommodations should be relevant to a student's needs.

Because CBT large-scale testing often carry high-stakes that impact numerous students' testing procedures, identifying successful practices that support EL students' linguistic and academic needs are necessary so ELs can demonstrate what they know and can do.

Validity of EL Test Accommodations

There have been multiple studies examining the validity of EL test accommodations. In an experimental study that compared non-EL students' performance with accommodations and non-EL students' performance without accommodations (Abedi, 2009; Abedi et al., 2020), the math performance of non-EL students were not statistically significantly different from those students receiving EL test accommodations (i.e., linguistically modification, English read aloud, and English glossary) (Abedi et al., 2020). However, a bilingual glossary improved non-EL accommodated students' math performance compared to non-EL without accommodations. Li and Suen (2012b) conducted a meta-analysis with 21 studies. They found that while EL test accommodations improve EL students' performance by a .16 standard deviation unit, the accommodations do not influence non-EL students' performance. Although there are mixed results regarding the validity of EL test accommodations, it is more likely that EL test accommodations do not cause unfair advantages to EL students.

Effectiveness of EL Test Accommodations

Despite the theoretical support of test accommodations for EL students, the research presents mixed results regarding test accommodations' effectiveness. EL test accommodations could be divided in two categories, dictionaries and linguistic modification (e.g., translation, linguistic simplification). Dictionaries may help reduce the linguistic complexity so EL students can understand unknown words. Bear in mind that defining the words may create a threat to the assessment's validity due to giving unfair advantages to EL students (Abedi, Courtney & Leon, 2003; Acosta et al., 2008). There was no statistically significant difference between EL students who used a dictionary and those who did not use a dictionary in reading tests (Albus et al., 2001). On the other hand, testing the effectiveness of pop-up glossaries with 4th and 8th grade students demonstrated that pop-up glossary test accommodation could increase ELs' test scores by a .50 standard deviation unit (Abedi, 2009). In addition, there are bilingual dictionaries and picture dictionaries that provide visual support for English words. The previous meta-analysis studies combined customized dictionaries, glossaries, bilingual dictionaries, picture dictionaries, and pop-up glossaries, and they did not find a statistically significant effect of dictionary and glossary accommodations (Kieffer et al., 2012; Liu & Suen, 2012a; Rios et al., 2020).

Translation and linguistically modified tests are also EL test accommodations. Translation accommodation can take different forms, such as taking the test in the native language or taking only the test instructions in the native language. Turkan and Oliveri (2014) articulated 12 out of 50 states offer translation accommodation to EL students, and they highlighted the effectiveness of translation accommodation depended on the quality of the test translation. There is limited research on native language accommodations (Kieffer et al., 2012), but it is important to match the language of assessment with the language of instruction (Abedi et

al., 2004; Kieffer et al., 2012). A combination of 12 studies' effect sizes demonstrates that linguistic simplification accommodation can improve ELs performance and decrease the achievement gap between ELs and non-ELs from 9% to 19% (Kieffer et al., 2012).

Computer-Based Testing for EL Students

In 2015, most K-8 students took a high-stakes state summative assessment on the computer; in fact, only 15% of the assessments were offered using a paper-based format (Strategies, 2015). CBT can simultaneously allow multiple accommodations, combining pop-up glossary and read-aloud accommodations (Russell et al., 2009). According to Abedi (2014), CBT allows effective test accommodations, which are not readily available for paper-based testing, so that it may provide an efficient assessment mode for EL students. Considering the importance of providing test accommodations and CBT's benefits for creating accessible and fair assessments, examining the validity and effectiveness of test accommodations on CBT becomes critical for an appropriate assessment process.

Comparing test accommodations, pop-up glossary on a computer, customized dictionary, extra time, and small-group testing reveals that computer testing and extra time effectively improve assessment accessibility to EL students without raising any validity concern (Abedi, 2009). Abedi et al. (2020) examined the effectiveness of EL test accommodations on CBT using an experimental design and surprisingly found that there were no significant gains for participants who used CBT accommodations, and in some cases, there was a negative impact. The study found there were statistically significant lower scores for EL participants who used Spanish math tests and bilingual glossary accommodations than EL students' who did not use any accommodations. Furthermore, there were no differences between the experimental and control groups for linguistic modification, English read-aloud, and English glossary.

A Meta-analysis of EL Test Accommodations

As EL test accommodation literature grew, there have been multiple meta-analyses to summarize EL test accommodations literature. Test accommodations can increase students with disabilities academic testing scores up to .16 standard deviation units, but out of 30 studies, only seven studies included EL students (Chiu & Pearson, 1999). Kieffer et al. (2009) conducted a meta-analysis examining the effectiveness of test accommodations for EL students on large-scale assessments and found that English dictionaries and glossaries have a small but statistically significant effect on ELs performance. In a meta-analysis by Pennock-Roman and Rivera (2011), which included 14 studies, simplified English test accommodation was more beneficial if EL students' language proficiencies are at an intermediate level. This meta-analysis also revealed that computer-administered glossaries are effective regardless of time restriction.

On the other hand, Li and Suen's (2012a) meta-analysis, including 19 studies, suggested English proficiency level and time restriction influence the effectiveness of the EL test accommodations. Rios et al. (2020) conducted a meta-analysis, which included 26 studies and 95 effect sizes, and concluded that accommodations could improve ELs test performance by .16 standard deviations. This meta-analysis also demonstrated that EL test accommodations are less effective for math/science content than non-math/science test contents.

The Rationale for the Current Study

Rios et al. (2020) highlighted the disparity between EL test accommodations research and practice. Similarly, previous meta-analyses (Kieffer et al., 2009; Li & Suen, 2012; Pennock-Roman & Rivera, 2011; Rios et al., 2020) combined paper-based test accommodations and CBT accommodations even though the majority of state accountability assessment is delivered on CBT. Thus, the validity and effectiveness of EL test accommodations for only CBT have not

been examined. Considering the expansion of CBT and disparity between EL test accommodation research and practices, it is necessary to summarize the research on EL test accommodation on CBT using a meta-analysis.

The purpose of this study is to investigate the validity and effectiveness of EL accommodations on CBT. This meta-analysis uses the random-effect model to quantify the average effects of EL test accommodations on CBT. The following research questions have guided this study:

1- Do EL test accommodations on CBT influence non-EL students' academic performance?
2- To what extent are CBT test accommodations effective in improving EL students' academic performance?
3- What factors influence the effectiveness of EL test accommodations?

Method

A meta-analysis approach, which aggregates quantitative research findings to uncover the patterns of the literature and build new theories, was introduced by Glass (1976; 1977). The general structure of conducting a meta-analysis includes a statement of the problem, literature search for relevant studies, quality evaluation, analyzing the outcomes, interpreting the evidence, and displaying the results (Cooper, 2017, p.25).

Literature Search

Electronic and manual literature searches were used to capture all the relevant research regarding EL test accommodations in CBT. An electronic literature search was conducted targeting the major databases, Education Resources Information Center (ERIC), Educational Administration Abstracts, Journal Storage (JSTOR), ProQuest, and PsycINFO. The electronic literature search was conducted in May 2020, and different combinations of the following

keywords were used to find related studies: *accommodations, test accommodations, English language learners, computer, computer-based assessment*s. Search results included peer-reviewed articles, technical reports, theses/dissertations, and conference proposals. Peer-reviewed journal articles and technical reports were included because of the quality of the research. Theses/dissertations and conference proposals were included to eliminate the publication bias because studies with statistically significant effects are more likely to be published (Glass, 1977). This electronic literature search for research was published between 1997 to 2020, and the location was limited to the United States.

Inclusion Criteria

The inclusion criteria used to select studies included (a) empirical quantitative studies that analyze CBT test accommodations for EL or bilingual students in K-12 settings, (b) studies that reported effect sizes or enough data to compute effect sizes, and (c) studies with at least one EL test accommodation on CBT. EL accommodation studies using a paper-based format were excluded (e.g., Abedi et al., 2001; Deysson, 2013; Fairbairn, 2006). Studies were excluded if they did not include test performance (i.e., Roohr & Sireci, 2017) or did not report enough information to compute effect sizes (i.e., Cohen et al., 2017).

In addition to published articles and technical reports, conference presentations, dissertations, and theses were included to eliminate the probability of publication bias. Borenstein et al. (2009) stated that comprehensive research is the ideal way to deal with publication bias so that multiple databases (ERIC, JSTOR, ProQuest, PsycINFO) were included in the literature search, and the time frame was from 1997 to 2020.

Data Coding

The following variables were included in data analysis based on the literature: grade, content, accommodations (Dictionary/Glossary, Linguistic Modification, Translation, Read-aloud), and the use of multiple accommodations. Like Li and Suen (2012a), the grade was coded as 0 for K-6 and 1 for K 7-12. It was hypothesized that test accommodations would be statistically more effective for K-6 EL students (Rios et al., 2020). Test content was coded as Reading and Math/Science as Li and Suen (2012) suggested. Because all eligible studies were in Math/Science content, this variable was not included in the analysis (see Table 1-2). Accommodations were dictionaries, including a bilingual glossary, pop-up glossary and picture dictionary, translation, linguistic simplification, and English read-aloud. Accommodations were categorized in three groups, dictionary, translation, and others due to small number of studies. Two dummy variables were coded for dictionary and translation accommodations, and others (i.e., linguistic simplification and read-aloud), with the others accommodation serving as the reference/comparison group of accommodations.

To examine the quality of the coding, a second reader independently coded over half of the studies. The average agreement between the two raters was 85%. The coders met to examine the disagreements and resolved all disagreements.

Data Analytical Procedure

Cohen's *d* effect size was used to compute the standardized mean difference between the treatment and control groups according to the following formula:

$$d = \frac{X_1 - X_2}{SD_{pooled}} \quad (1)$$

X_1 is the mean of the treatment group (receiving accommodations), X_2 is the control group (no accommodations) mean, and SD_{pooled} is the pooled standard deviation. Since Cohen's *d* tend to

overestimate effect sizes from small samples, Cohen's *d* was converted to Hedges' *g* by using the following formulas:

$$g = j * d \qquad and \qquad j = 1 - \frac{3}{4(n_1 + n_2 - 2) - 1} \tag{2}$$

In this formula, g is Hedges' g, d is Cohen's d, n_1 is the sample size of the treatment group, and n_2 is the control group's sample size.

All the analyses were conducted in R software with Meta and Metafor packages. Outliers were examined before computing the average effect size. Two studies (three effect sizes) were detected as outliers. The analyses were conducted with outliers and without outliers; we did not remove the outliers because of minor differences between the results. The random-effect model was used to compute the average effect size and effect size heterogeneity for EL test accommodations' validity and the effectiveness of EL test accommodations on CBT. Heterogeneity of effect size and the average effect size estimates were computed with a random-effects (intercept-only) model according to restricted maximum likelihood estimation. Inverse variance weight was applied to the effect sizes. I^2 statistics were used to measure heterogeneity where I^2 *<50%* indicates small heterogeneity, 50%< I^2<75% indicates medium heterogeneity, and I^2 >75% indicates large heterogeneity (Higgins, & Thompson, 2002).

The moderator model for the effectiveness of EL test accommodations on CBT was conducted via restricted maximum likelihood estimation as follows:

$$Effectiveness_{Hedges' g} = \beta_0 + \beta_1 * (Grade) + \beta_2 * (Dictionary) + \beta_3 * (Translation) + e \tag{3}$$

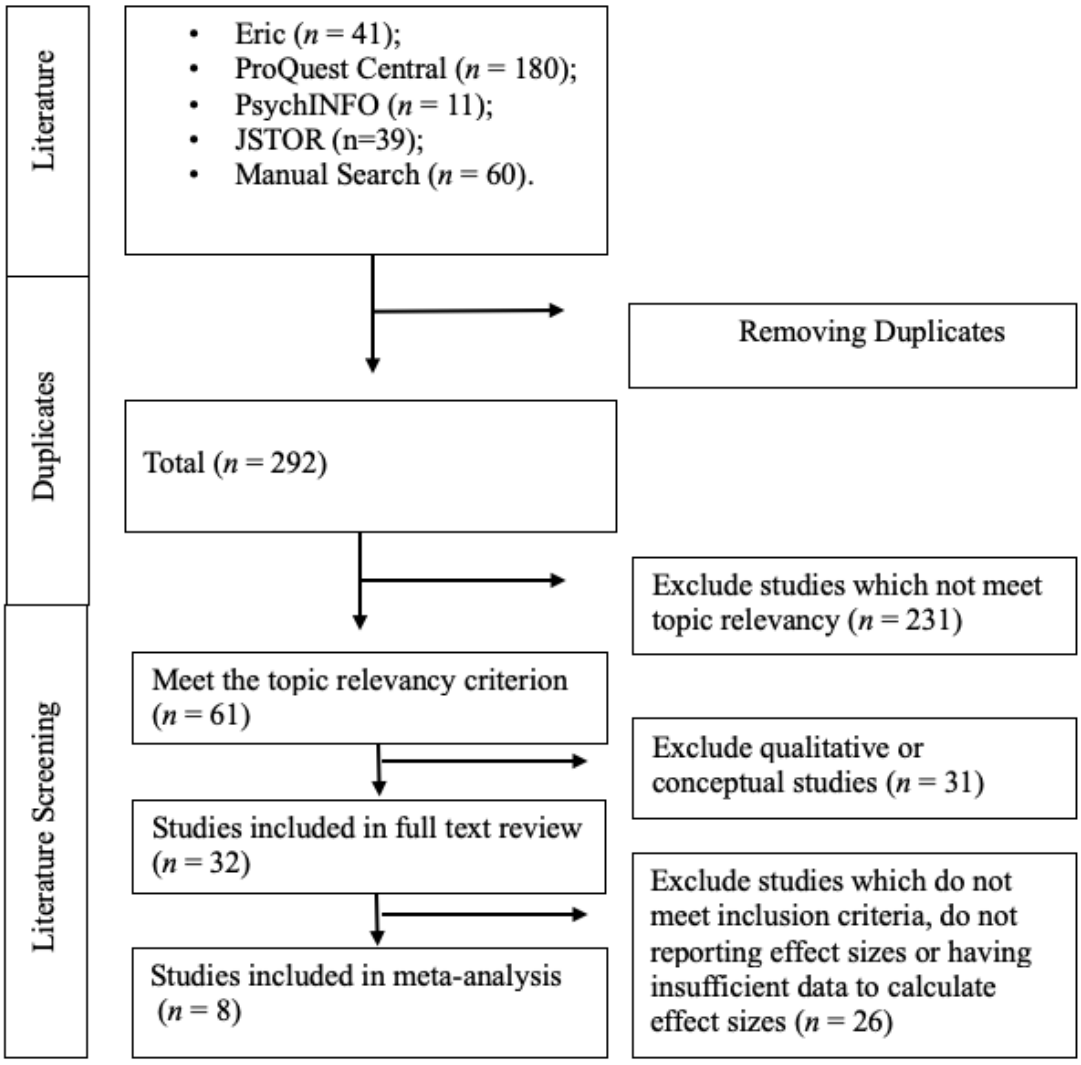
In this formula, β_0 represents the average effect size after controlling for the included variables, and *e* was the residual term.

Results

The PRISMA Flow Diagram for the literature search is shown in Figure 1-1. Each study's abstract was examined to understand if the study was empirical research about CBT EL test accommodations. In addition, previous EL test accommodation meta-analyses (Kieffer et al., 2009; Li & Suen 2012a, 2012b; Pennock-Roman & Rivera, 2011; Rios et al. 2020) references were reviewed for potential research that met the inclusion criteria.

Figure 1-1

PRISMA Flow Diagram of Literature Search and Screening



After literature screening, out of 292 studies, 61 studies met the topic relevancy criterion. Out of these, 32 of studies were eligible for full-text screening, and finally, eight studies were included in the meta-analysis. Only one of the eight eligible studies (12.5%) came from unpublished research.

The eight studies were conducted between 1999 and 2020. Although the sample size is small in this study, 25% of the studies were not included in the previous meta-analyses (Abedi et al., 2020; Ardasheva et al., 2018). Also, it is important to note that researchers used paper-based terminology regarding dictionary accommodations for ELs. We presented the description of each accommodation in Table 1-2 based on the researchers' choice. However, all dictionary accommodations were grouped as one category in the data analysis process.

Table 1- 2

The Summary of Eligible Studies

Study	Sample	Grade	Content	Accommodations	The implementation
Abedi (2009)	1149	4 & 8	Math	Pop-up Glossary	This study used PPT and CBT. Pop-up dictionary demonstrates a simple explanation of a word with the touch of the mouse.
Abedi et al. (2020)	1530	9	Math	Linguistic simplification, English read-aloud, English glossary, Translation, Bilingual glossary	Linguistic simplification reduces language demand without threating validity of the assessment. Read-aloud have audio files for questions so students can listen it. English glossaries include definition of selected words. Students took Spanish Glossary or Spanish Math (translation). Test if they were proficient in Spanish as getting 89 or higher in Spanish TIMER.
Kopriva et al. (2007)	272	3 & 4	Math	Picture dictionary, bilingual dictionary, and	Picture dictionaries present a picture of selected words with a mouse click. Bilingual dictionaries provide Spanish

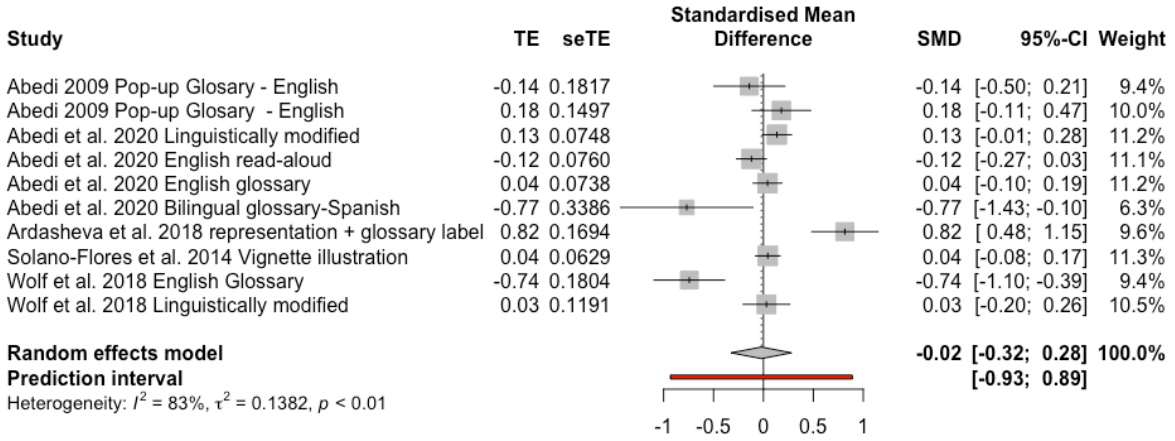
				English read-aloud, and combination of these	translation of the selected words. Students' Spanish proficiency was determined by their teachers. Read-aloud would read the items to students.
Ardasheva et al. (2018)	174	7	Science	Visual support + glossary	This accommodation provides visual representations of the words trough Google images and short definition of the words.
Robinson (2010)	3273	K & 1	Math & literacy	Translation	Spanish version of the Math and Literacy tests.
Solano-Flores et al. (2014)	728	8	Science	Illustration (Visual support)	This accommodation adds an illustration to the items that included only text.
Alt et al. (2013)	21	2	Math	Translation	Spanish version of the test
Wolf et al. (2018)	513	8 & 9	Math	English Glossary, Linguistic simplification	English glossaries provide a short definition of the words when students click the word. Linguistic simplification provides lexical and syntactic support.

The Validity of EL Test Accommodations on CBT

Five studies examined the validity of EL test accommodations on CBT, and these studies included 10 effect sizes and 2779 non-EL students in total. In Figure 1- 2, the random-effect model results for the validity of EL test accommodations were presented. The total effect size for the validity of EL test accommodations on CBT was -0.02 *SD* (*SE*=0.13; 95% CI: -0.3246, 0.2825; *p*=.88). Although there is a large heterogeneity (I^2=83%; *Q*=53.23 $p < .001$), a moderator analysis was not conducted due to the small sample size.

Figure 1-2

The Forest Plot of the Effect Sizes for the Validity of EL Test Accommodations on CBT



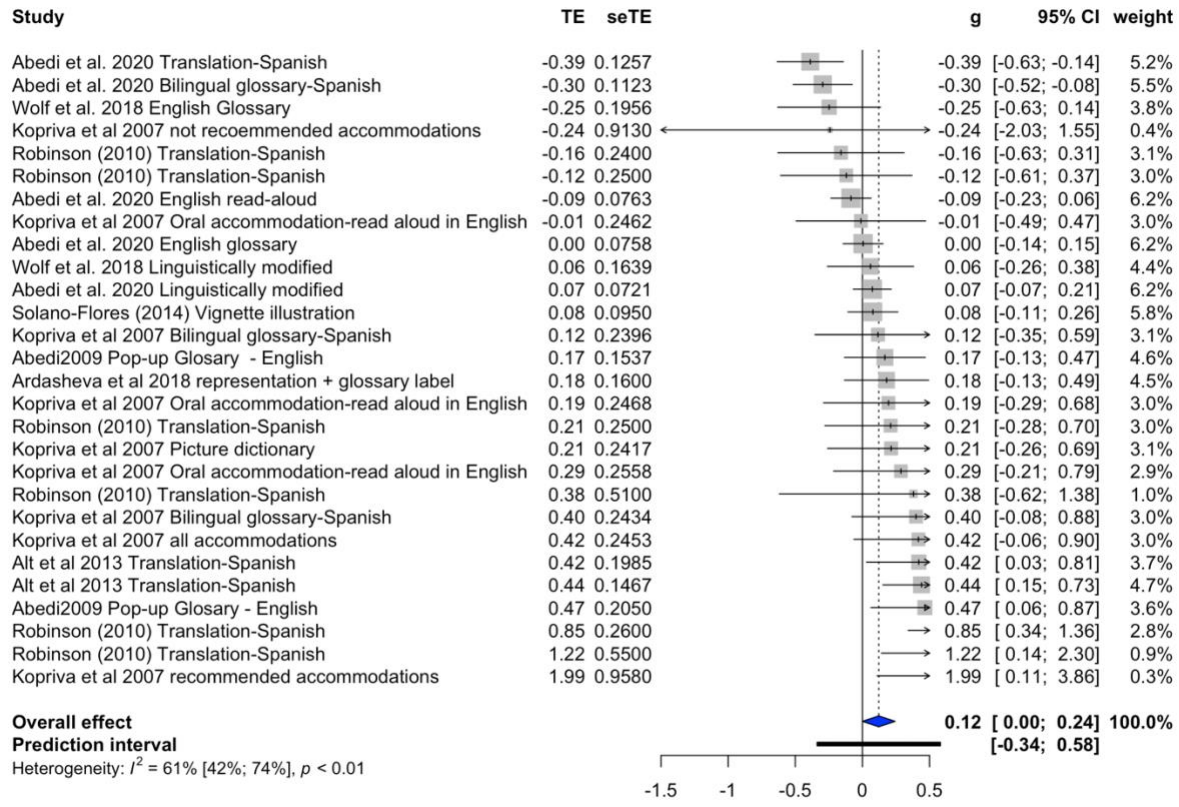
Note. In the forest plot, the center of the box represents the size of the treatment effect, and the black line shows the confidence interval. While a diamond shows the summary of the treatment effect, the confidence interval is represented with right and left extremes.

Effectiveness of EL Test Accommodations on CBT

In total, eight studies examined the effectiveness of EL test accommodations on CBT, and these studies included 28 effect sizes and 5987 EL students. Only two effect sizes (9%) came from unpublished literature. In Figure 1-3, the random-effect model results for the effectiveness of EL accommodations on CBT were presented. The total effect size of EL test accommodations on CBT was .12 *SD* (*SE*=0.06; 95% CI: 0.0002, 0.2433; p <.05). A moderator analysis was conducted because of the heterogeneity (I^2=61%; Q=69.86 p < .01).

Figure 1-3

The Forest Plot of the Effect Sizes for the Effectiveness of EL Test Accommodations on CBT



Note. In the forest plot, the center of the box represents the size of the treatment effect, and the black line shows the confidence interval. While diamond shows the summary of the treatment effect, the confidence interval is represented with right and left extremes.

According to 28 effect sizes, grade, dictionary, and translation variables were included in the model. Table 1-3 presents the model results. Four out of 8 studies included at least one K-6 sample, which was about 65% of the effect sizes. The results demonstrated a statistically significant difference in accommodation effectiveness when comparing samples in K-6 and Grade 7-12 *(β*= -0.43; $p < .001$), which indicated that accommodations' effectiveness was .43 SD lower in grade 7-12 compared to K-6. The estimated effect size for K-6 was .34 when the average values of the dictionary and translation variables were added to the regression equation,

while the estimated effect size for grades 7-12 was -.09. Although dictionaries, including bilingual glossaries and picture dictionaries, were the most common accommodations and accounted for 46% of the effect sizes, there was no statistically significant difference between a dictionary and other accommodations ((β= -0.124 p = .30). Similarly, translation accommodation (26%) did not have a statistically significant effect on EL test performance compared to other test accommodations (β= -0.201; p = .20).

Table 1-3

The Moderator Analysis for EL Test Accommodations

	Moderator Model (k=8, n=28) I^2 =37.45; τ^2= 0.016 (SE=0.0131)			
Moderator	Estimate	SE.	95% CI	*p*
Intercept	0.462	0.134	0.174, 0.750	0.003**
Grade	-0.431	0.113	-0.662, -0.200	0.001***
Dictionary	-0.124	0.117	-0.365, 0.117	0.298
Translation	-0.201	0.151	-0.512, 0.110	0.195

Note. *p< .05; **p< .01; ***p< .001.

Discussion

The findings of this study are similar to the paper-based accommodations for ELs. Results from this study suggest that non-ELs' performance was not improved by CBT accommodations (-.02 *SD*), which is consistent with Li and Suen's (2012b) meta-analysis. The evidence indicates that EL test accommodations on CBT do not provide unfair advantages to EL students.

This meta-analysis suggested that EL test accommodations on CBT have a small and statistically significant effect on improving EL students' academic performance. EL students with accommodations on CBT received .12 SD higher scores than EL students without test accommodations on CBT. Rios et al. (2020) meta-analysis found the average effect of test

accommodations as .16 SD, which is the between confidence interval of this study. Even though Rios et al. (2020) and Li and Suen (2012a) meta-analyses did not report a statistically significant difference between grade levels, this study showed that EL test accommodations on CBT were more effective for K-6 compared to grades 7-12. The reasons for these differences are not clear, and future studies should focus on how the use of accommodations and other factors (e.g., English language proficiency level) may vary across grade levels.

Although the dictionary is a common EL test accommodation, the results did not demonstrate a statistically significant effect of dictionary accommodation on CBT. This result is similar to Rios et al.'s (2020) and Li and Suen's (2012a) meta-analysis results. Small sample size could be a potential reason for the statistically insignificant result, while limited evidence is available whether dictionaries increase content accessibility (Rios et al., 2020). Possibly, lack of alignment between the language of instruction, including textbooks and teachers' instructional language, and EL test accommodations could be the reason for the statistically insignificant effect of EL test accommodations (Abedi et al., 2020).

This meta-analysis provides some evidence about the validity and effectiveness of EL test accommodations on CBT. However, there are some limitations. First, the sample size of the meta-analysis was small. Even though CBT is a common practice for states' high-stakes academic testing, there is limited research that examines EL accommodations on CBT. Second, the small sample size could be the reason for the statistically insignificant effect of dictionary and translation test accommodation. Third, the EL population is a heterogeneous group of students, and accommodations should be provided based on an individual student's needs. Considering student needs in providing test accommodations important, but this meta-analysis could not examine the effect of the EL population's diversity in terms of language proficiency

level and student background information because the studies did not provide enough background information on participated EL students. Lastly, the quality of a meta-analysis based on the studies included. Thus, we acknowledged that these eight studies were different in terms of sample size, grades, the number of accommodations used, types of accommodations, and how these accommodations were implemented.

Future Studies

As the use of technology increases in educational settings, CBT will become the common assessment practice. The results of this study are important because the majority of states' large-scale assessments are administered via CBT. This meta-analysis provides some evidence that suggests EL test accommodations on CBT have a small effect on improving EL students' academic performance. In addition, this study revealed the need for more experimental studies examining CBT accommodations. CBT allows the seamless integration of accommodations into the testing process and permits gathering extensive data, such as response time and the frequency of accommodation use (Roohr & Sireci, 2017). Therefore, examining the effect of the frequency of accommodations on ELs academic achievement is a much-needed research area.

While Rios et al.'s (2020) meta-analysis included 26 experimental design studies about EL test accommodations, there were only eight eligible studies for this meta-analysis. This study agreed with Rios et al.'s findings regarding EL test accommodation having limited evidence to support the effectiveness of accommodation. All studies in this meta-analysis were in math and science content areas, so future EL test accommodation research should include other content areas. Since this study indicated that EL test accommodations are more effective in K-6 grades, future research should examine which accommodations are effective for different grade levels. ELs are a diverse group of students, which requires additional research to examine effective

accommodations based on students' English language proficiency and other student characteristics. Since the benefits of accommodations are dependent on students' backgrounds and needs (Abedi, 2013), examining the moderating effects of these factors could provide insight into increasing the effectiveness of accommodations. Until additional empirical studies are conducted examining student characteristics, specific types of accommodations, and specific content areas, we will not fully understand the impact of CBT accommodations and potential methods for improving fairness and accessibility for all students.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*(1), 4–14.

Abedi, J. (2006a). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record, 108*(11), 2282.

Abedi, J. (2006b). Language issues in item-development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Erlbaum.

*Abedi, J. (2009). Computer testing as a form of accommodation for English language learners. *Educational Assessment, 14*(3-4), 195-211.

Abedi, J. (2013). Validity issues in designing accommodations for English language learners. In *The Routledge handbook of language testing* (pp. 62-76). Routledge. Accessed on: 16 Jan 2020 https://www.routledgehandbooks.com/doi/10.4324/9780203181287.ch3

Abedi, J. (2014). English language learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology,* 10(2), 1-30.

Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Tech. Rep. No. 586). Los Angeles, CA: University of California, National Center for Research on Evaluation Standards and Student Testing.

Abedi, J., & Levine, H. G. (2013). Fairness in assessment of English learners. *Leadership*, *42*(3), 26–38. https://eric.ed.gov/?id=EJ1004517

*Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence‐based

computer accommodation decision system. *Educational Measurement: Issues and Practice*.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Abedi, J., Lord, C., Kim, C., & Miyoshi, J. (2001). *The effects of accommodations on the assessment of LEP students in NAEP* (CSE Tech. Rep. No. 537). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Acosta, B. D., Rivera, C., & Willner, L. S. (2008). *Best practices in state assessment policies for accommodating English language learners: A delphi study*. George Washington University Center for Equity and Excellence in Education.

Albus, D., Bielinski, J., Thurlow, M., & Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test* (LEP Projects Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Albus, D., Thurlow, M., Liu, K., & Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research, 98*(4), 245-256.

*Alt, M., Arizmendi, G. D., Beal, C. R., & Hurtado, J. S. (2013). The effect of test translation on the performance of second grade English learners on the KeyMath-3. *Psychology in the Schools*, 50, 27-36.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014). *Standards for educational and*

psychological testing. Washington, DC: American Educational Research Association, American Psychological Association and National Council on Measurement in Education.

*Ardasheva, Y., Wang, Z., Roo, A. K., Adesope, O. O., & Morrison, J. A. (2018). Representation visuals' impacts on science interest and reading comprehension of adolescent English learners. *Journal of Educational Research,* 111(5), 631–643. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1182786&site=ehost-live&scope=site

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex, UK: John Wiley & Sons.

Chiu, C. W. T., & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommodations for special education and limited English proficiency students*. Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT.

Cohen, D., Tracy, R., & Cohen, J. (2017). On the effectiveness of pop-up English language glossary accommodations for EL students in large-scale assessments. *Applied Measurement in Education*, 30(4), 259-272.

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach* (Vol. 2). Sage publications.

Deysson, S. L. (2013). *Equity for limited English proficient students regarding assessment and effectiveness of testing accommodations: A study of third graders* (Order No. 3550425). [Doctoral Dissertation, The George Washington University]. Available from Education Database; ProQuest Central. (1287056937). Retrieved from https://search-proquest-com.librarylink.uncc.edu/docview/1287056937?accountid=14605

Dolan, R. P., Burling, K. S., Rose, D., Beck, R., Murray, E., Strangman, N., Jude, J., Harms, M., Way, W. D., Hanna, E., Strain-Seymour, E., & Nichols, A. (2010). *Universal Design for Computer-Based Testing (UD-CBT) guidelines*. Iowa City, IA: Pearson.

Every Student Succeeds Act. (2015). *Every Student Succeeds Act of 2015*, Pub. L. No 114 – 95 (2015 - 2016).

Fairbairn, M. B. (2006). *English language learners' performance on modified science test item formats: A pilot study*. (Unpublished doctoral dissertation, University of Iowa, IA.)

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved from http://centeroninstruction.org/files/ELL3-Assessments.pdf

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher,* 5, 3–8.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351–379.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention, 29*(3), 35-45.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201.

Kieffer, M. J., Rivera, M., & Francis, D. J. (2012). Based recommendations for the use of accommodations in large-scale assessments: 2012 Update. *Practical Guidelines for the Education of English Language Learners*. Book 4. Center on Instruction.

*Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, *26*(3), 11-20.

Li, H., & Suen, H. K. (2012a). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education*, 25(4), 327-346.

Li, H., & Suen, H. K. (2012b). Are test accommodations for English language learners fair? *Language Assessment Quarterly*, 9(3), 293-309.

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., and Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017- 144). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., and Bullock Mann, F. (2018). *The Condition of Education 2018* (NCES 2018-144). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144.

Miley, S. K., & Farmer, A. (2017). English Language Proficiency and Content Assessment Performance: A Comparison of English Learners and Native English Speakers Achievement. *English Language Teaching, 10*(9), 198–207.

National Assessment of Educational Progress (NAEP) (2017). *Academic performance and outcomes for English learners: Performance on national assessments and on-time graduation rates*. Washington, DC: US Department of Education. https://www2.ed.gov/datastory/el-outcomes/index.html

National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs. (2007). Retrieved from http://sbo.nn.k12.va.us/esl/documents/ ncela_fast_faqs.pdfPennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*(3), 10-28.

Polat, N., Zarecky-Hodge, A., & Schreiber, J. B. (2016). Academic growth trajectories of ELLs in NAEP data: The case of fourth- and eighth-grade ELLs and non-ELLs on mathematics and reading tests. *Journal of Educational Research, 109*(5), 541-553. doi:10.1080/00220671.2014.993461

Rios, J. A., Ihlenfeldt, S. D., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence‐based? A multistudy systematic review and meta‐analysis. *Educational Measurement: Issues and Practice*.

Rivera, C. (2003). State assessment policies for English language learners. Paper presented at the Presented at the 2003 Large-Scale Assessment Conference.

*Robinson, J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher*, 39(8), 582-590.

Roohr, K. C., & Sireci, S. G. (2017). Evaluating computer-based test accommodations for English Learners. *Educational Assessment,* 22(1), 35-53.

Russell, M., Hoffmann, T., & Higgins, J. (2009). Meeting the needs of all students: A universal design approach to computer-based testing. *Innovate: Journal of Online Education, 5*(4).

Shafer Willner, L. S., Rivera, C., & Acosta, B. D. (2008). *Descriptive study of state assessment policies for accommodating English Language learner*s. Arlington, VA; The George Washington University Center for Equity and Excellence in Education. Retrieved from http://ceee.gwu.edu

Shepard, L. A., Taylor, G., & Betebenner, D. W. (1998). *Inclusion of limited-English-proficient students in Rhode Island's grade 4 mathematics performance assessment.* Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3–13.

*Solano-Flores, G., Wang, C., Kachchaf, R., Soltero-Gonzalez, L., & Nguyen-Le, K. (2014). Developing testing accommodations for English language learners: Illustrations as visual supports for item accessibility. *Educational Assessment, 19*(4), 267-283.

Stone, E. A. & Cook, L. L. (2018). Fair testing and the role of accessibility. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible instruction and testing practices,* Second edition, (pp. 59-73). Springer.

Strategies, E. (2015). *Pencils down: The shift to online & computer-based testing*.

Thurlow, M., Laitusis, C. C., Dillon, D. R., Cook, L. L., Moen, R. E., Abedi, J., & O'Brien, D. G. (2009). *Accessibility principles for reading assessments*. Minneapolis, MN: National Accessible Reading Assessments Projects.

Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). Computer-based testing: Practices and considerations (Synthesis Report 78). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Turkan, S., & Oliveri, M. E. (2014). *Considerations for providing test translation accommodations to English language learners on common core standards‐based assessments*. ETS Research Report Series, 2014(1), 1-13.

United States Department of Education (2015). *Every Student Succeeds Act*. Retrieved from https://www.ed.gov/essa?src=rn

Wolf, M. K., Herman, J. L., Kim, J., Abedi, J., Leon, S., Griffin, N., et al. (2008). *Providing validity evidence to improve the assessment of English language learners* (CSE Tech. Rep. No. 738). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

*Wolf, M. K., Yoo, H., & Guzman-Orth, D. (2018, April). Exploring process data to understand English learners' assessment performance: Response time and glossary use. Paper presented at the annual meeting of the National Council on Measurement in Education, New York City, NY.

CHAPTER 2 [ARTICLE II]: PREDICTING POSTSCHOOL OUTCOMES FOR ENGLISH LEARNERS WITH DISABILITIES: SECONDARY ANALYSIS OF DATA FROM THE NATIONAL LONGITUDINAL TRANSITION STUDY 2

Gezer, T., Flowers, C., & Mazzotti, V. Predicting postschool outcomes for English learners with disabilities: Secondary analysis of data from the National Longitudinal Transition Study 2. *Career Development and Transition for Exceptional Individuals*

Employment, postsecondary education, and independent living are the common indicators of a successful postschool transition for students with disabilities (SWD; Mazzotti et al., 2021; Mazzotti et al., 2016; Test, Mazzotti et al., 2009). Researchers in secondary transition have identified evidence-based practices that SWD develop and maintain skills that prepare them for postsecondary education, employment, and independent living (e.g., McDowell, 2004; Rowe et al., 2021; Test, Fowler et al., 2009). Other researchers, using correlational research, have identified predictors of positive postschool outcomes (PSO) for SWD (e.g., Mazzotti et al., 2021; Roessler et al., 1990).

English learners (EL) with disabilities are a growing population as the number of ELs increases in the United States. Approximately 12% of students are SWD, and nearly 10% of SWD have limited English proficient (Lipscomb et al., 2017). The Individuals with Disabilities Education Act (IDEA, 2004) granted SWD the right to have access to free and appropriate public education, and a significant amount of literature is available regarding instruction, needs, and transitions for SWD (Artiles & Klingner, 2006; Daviso et al., 2011; Powers et al., 2005). However, there is limited research on the postschool outcomes of EL with disabilities. This study fills the literature gap by examining postschool outcomes for EL with disabilities using data from the National Longitudinal Transition Study-2 (NLTS2).

Literature Review

Postschool Outcomes

PSO provides a picture of life after high school for former special education students concerning postsecondary education, employment, and independent living (Center for Change in Transition Services, 2021). Special education research has been examining SWD's PSO and transition for over 30 years. Recent systematic literature reviews focus on employment, postsecondary education, and independent living as the indicators of a successful postschool transition for SWD (Mazzotti et al., 2021; Mazzotti et al., 2016; Test, Mazzotti, et al., 2009).

For many SWD, employment is a primary focus of postschool transition (Cameto et al., 2004). Even though EL and EL with disabilities may access education regardless of citizenship, or immigration status, employment can be problematic for them (Trainor et al., 2014). Comparing employment rates for jobs outside the home since high school indicates a difference between EL with disabilities who got a job outside of the home (80%) and SWD (not EL; 91%) who become an employee outside of the home (Trainor et al., 2016).

Postsecondary education often leads to better job opportunities and higher salaries so that many high school students aim to enroll in postsecondary education (Marcotte et al., 2005). High school graduation and postsecondary education enrollment outcomes are similar for EL with disabilities and SWD (Trainor et al., 2016). They concluded that nearly 66% of EL with disabilities graduated from high school, and 61% of them continued postsecondary education based on secondary analysis of NLTS 2.

Independent living is defined as living alone, with a spouse, partner, or roommate (Newman et al., 2011). Research demonstrated that independent living outcomes vary based on students' disability. The percentage of independent living for some of the disability categories

are as follows learning disability (64.9%), emotional disturbance (63.1%), other health impairment (58.2%), visual impairment (55.4%), intellectual disability (36.3%), deaf-blindness (26.4%), and autism (17%; Newman et al., 2011).

Factors Associated with PSO

While Kohler (1996) reported that parent involvement, future planning, and school programs are critical aspects of postschool transition, Trainor et al. (2020) described the postschool transition in four layers which are culture (family, community, social capital), services & supports (assessment, planning instruction etc.), levers (laws, funding, policies) and quality of life (experiences and outcomes). According to systematic reviews of NLTS2 studies, some of the predictors of postschool success for SWD include career awareness, parent involvement, high school diploma status, independent living skills, inclusion in general education, work experience/paid employment, work-study, vocational education, and social skills (Mazzotti et al. 2021; Mazzotti et al., 2016; Test, Mazzotti et al., 2009). In addition to these predictors, parent expectations, goal setting, decision making/youth autonomy, and travel skills are the predictors of postschool success (Mazzotti et al., 2016). Prior studies (Dell'Armo & Tassé, 2019; Kirby, 2016) examined demographic information, adaptive behavior, and parent expectations related to postschool transition outcomes. This study was built on Dell'Armo and Tassé's (2019) and Kirby's (2016) studies, and predictor latent variables were categorized as transition planning, adaptive behaviors, and parent expectations. Table 2-1 presents the summary literature on postschool transition predictors. The following sections described each category.

Table 2-1

Literature Summary of PSO Predictors

Latent Variables	Variables	Literature support
Transition planning	Transition goals, Youth involvement	Mazzotti et al. (2016), Trainor et al. (2019), Wei et al. (2016), Newman & Madaus (2015), Park & Bouck (2018)
Adaptive behaviors	Selfcare skills, Functional mental skills, Social skills	Dell'Armo and Tassé (2019), Carter et al., (2012), Papay & Bambara (2014), Mazzotti et al. (2016), Shattuck et al. (2012); Wei et al. (2017)
Parent expectations	Parent expectations of PSO: Postsecondary education, employment, and independent living.	Wagner et al., (2014), Kirby (2016), Dell'Armo and Tassé (2019), Trainor et al. (2019), Mazzotti et al. (2021), Rojewski et al. (2014)

Transition Planning

Transition planning is a significant predictor of postschool transition (Mazzotti et al., 2016; Mazzotti et al., 2021). Receiving transition planning supports SWD to evaluate their strategies and choices (Newman et al., 2016; Newman & Madaus, 2015). While the majority of EL with disabilities (92.9%) received transition planning, only half of their transitional goals included postsecondary education (42%), competitive employment (58%), and independent living (53%; Trainor et al., 2016). Schools play an important role in developing EL with disabilities postschool transition goals. According to Trainor et al. (2019), 51% of EL with disabilities indicated that most schools came up with transition goals, but this ratio is significantly low for SWD (39%).

Although EL with disabilities and SWD transition programs are similar in many aspects, EL with disabilities only benefit from transition programs based on their language proficiency in

addition to their disabilities (Hoover & Patton, 2005; Newman et al., 2016). Also, Trainor et al. (2014) suggested culturally responsive methods, where educators interact with families to understand students' characteristics, for transition education to eliminate bias. Similarly, EL with disabilities and their parents may have a different immigration status, which can cause anxiety beyond the cost of postsecondary education, and educators should provide information about the transition that is sensitive to citizenship status (Trainor et al., 2019).

Adaptive Behaviors

Adaptive behaviors consist of practical social and conceptual skills, which help maintain daily activities (Tassé et al., 2012). Self-care and independent living skills are potential predictors for employment (Carter et al., 2012; Mazzotti et al., 2016; Mazzotti et al., 2021) and postsecondary education (Mazzotti et al., 2016; Mazzotti et al., 2021; Papay & Bambara, 2014). Social skills were defined as “behaviors and attitudes that facilitate communication and cooperation” (Rowe et al., 2015, p.122). According to Mazzotti et al.’s (2016; 2021) systematic reviews, social skills are potential predictors of employment and postschool education for SWD.

Parent Expectations

Parent expectations is a research-based predictor of postschool employment (Mazzotti et al., 2021) and a significant predictor of postsecondary education (Wagner et al., 2014). EL with disabilities and their parents share similar expectations about postsecondary education according to NLTS 2012 data (Trainor et al., 2019). While 33% of EL parents with disabilities did not expect their child to continue postsecondary education, 25% of EL with disabilities' parents expected their student to pursue a 4-year college degree. According to the same study, EL with disabilities' parents have lower postsecondary education expectations than SWD's parent expectations, although the difference in parent expectations was insignificant. On the other hand,

both student groups' expectations are significantly lower than general education students' and their parents' expectations. For instance, 41% of general education students were expected to attain a 4-year college compared to SWD (31%) and EL with disabilities (25%). Although a similar percentage (58%) of EL with a disability and SWD parents attended the transition planning meeting, EL with disabilities' parents (49%) were less likely to talk about postschool education and career opportunities than SWD parents (60%; Trainor et al., 2019).

English Language Learners

Diverse demographic characteristics, including language, immigration, poverty, and disabilities, may be barriers to providing equal educational access to EL (Artiles & Klingner, 2006). The proportion of EL with disabilities is 14% of EL students in the US (McFarland et al., 2017). The majority of EL with disabilities is Latino (71%) and lives below the state poverty level (85%), and these proportions are significantly higher than SWD and general education student populations (Trainor et al., 2019). The same study shows that approximately 33% of EL with disabilities' parents did not graduate from high school compared to 12% of SWD and general education' parents. The majority of ELs are born in the United States, and some arrived in the US as immigrants (Flores et al., 2017). This mobility may cause the interruption of EL students' academic progress (Calderón et al., 2011). Besides, some school districts could enforce a formal/informal de facto policy, which does not allow EL with disabilities to receive language and special education services even though this is against the federal law (US Department of Justice & US Department of Education, 2015).

Similar to SWD, EL students have to overcome unique challenges in order to succeed in school. For instance, regardless of age and teaching experiences, teachers do not support rigorous instruction for EL students as they do for general education students, where critical

thinking was the indicator of rigorous instruction (Murphy & Torff, 2019). This may cause additional problems for EL students in postschool transition. For instance, many EL students have difficulties meeting college qualifications impacting their ability to apply for a four-year college (Kanno & Cromley, 2015). Roessingh and Douglas (2012) examined EL students' transition to college and found that EL was not ready for college's high literacy demands. Similarly, Watkins (2015) examined seven Latino EL students' transition from high school to a community college using phenological experiences. The author indicated that EL students might not understand the transition program's importance, and they need detailed information about the transition process.

Peers, parents, and school personnel are crucial components of social supports. Families, peers, and teachers are a source of support during EL postschool transition (Watkins, 2015). Baker (2017), however, demonstrated that the source of social support is less important than the kind of support EL students receive related to motivation, the course selection, and academic contents exemplify emotional, informational, and instrumental support, respectively.

Research Purpose

An extensive amount of research exists regarding SWD transition programs (Greene, 2014) and EL postschool transition (Roessingh & Douglas, 2012; Schlaman, 2019), but little is known about specific needs related to the transition supports EL with disabilities need in high school to help them attain positive PSO (Trainor et al., 2016; Wanzek et al., 2016). Although some research is available about the characteristics of EL with disabilities (Trainor et al., 2016) and the transition to postsecondary education (Trainor et al., 2019), this study aims to examine the direct and indirect relationships between EL with disabilities PSO and transition planning, family factors, adaptive behaviors using structural equation modeling (SEM). $H_{1:}$ Transition

mediates the positive effect of adaptive behavior and parent expectations on PSO. H_2: There is an indirect effect of adaptive behavior on transition planning and PSO. H_3: There is an indirect effect of parent expectations on PSO. H_4: There is a direct effect of parent expectation on transition planning. H_5: There is direct effect of transition planning on PSO.

1- To what degree do transition planning, family factors and adaptive behaviors predict postschool outcomes for EL with disabilities?

Method

This study examined predicting variables of a successful postschool transition for EL with disabilities by using the restricted data from the second iteration of the National Longitudinal Transition Studies, the NLTS2, which was conducted between (2001-2009) with five waves of data collection (Newman et al., 2011). Participants' ages were 21-25 years by the end of the NLTS2. This longitudinal study provided nationally representative data regarding SWD high school experiences and postschool transition.

NLTS2 used a two-stage stratified sampling design by randomly sampling 3,634 local educational agencies (LEA) and selecting 11,000 students from 12 disability categories recognized by IDEA 2004. These are intellectual disability, autism, deaf-blindness, hearing impairment (which includes deafness), emotional disturbance, orthopedic impairment, visual impairment, multiple disabilities, other health impairment, traumatic brain injury-specific learning disability, and speech or language impairment (Newman et al., 2009). The US Department of Education mandates rounding the unweighted sample size to the nearest 10 for restricted-use data. This research was approved by the University of North Carolina at Charlotte Institutional Review Board.

Sample for the Current Study

School program surveys Wave 1-2 and high school transcriptions from NLTS2 were used for EL with disabilities identification. Students were identified as EL if they met the following criteria: (a) according to school program surveys, students have taken bilingual education or instruction for EL; (b) students' high school transcriptions had listed hours for ESL/English; and (c) excluding students who only use sign language for communication or does not speak at all unless the student's transcript included ESL/English hours. Based on these criteria, there were 400 EL with disabilities in the NLTS2 sample. Among these students, 240 students, who had at least one PSO, postsecondary education, competitive employment, and independent living, were included in further analysis.

Gender, income, education level for the head of household, race/ethnicity, age, disability categories, grade, and PSO are displayed in Table 2-2. Most of the students were male (n = 150, 62%), reported their ethnicity as either Hispanic (n = 100, 42%) or White (n = 80, 33%), and were distributed across all disability categories and grade levels. Approximately half of the reported household income was $25,000 or less per year. The household's educational level tended to be at high school or GED (n = 80, 33%) or less than high school (n = 70, 33%).

Concerning PSO results, approximately 39% of the students attend postsecondary education, which was much lower than previous findings (Trainor et al., 2016). While 47% of them were competitively employed, 53% of them were not employed. According to Trainor et al. (2016), 80 % of EL with disabilities were employed outside the home since higher school, and this ratio is different from competitive employment, which is among the outcome variables in this study. Only 15% of them living independently, whereas 85% did not live independently.

Approximately 14% of young adults with intellectual disabilities lived independently (Dell'Armo & Tassé, 2019).

Table 2-2

Demographic Characteristics of Unweighted Sample

Variables	Frequency	%
Demographic Variables		
Gender		
• Male	150	62
• Female	90	38
Household Income		
• $25,000 and under	110	50
• $25,001-$50,000	60	27
• Over $50,000	40	18
Education Level for Head of Household		
• Less than high school	70	33
• High school or GED	80	38
• Some college	40	19
• Bachelor's degree or higher	30	14
Race/Ethnicity		
• White	80	33
• African/American	40	17
• Hispanic	100	42
• Asian/Pacific Islander	20	8
• American Indian/Alaska Native	<10	<4
• Multiracial/other	<10	<4
Age at Wave 1 – Mean (SD)	15.09 (1.17)	
Disability Categories		
• Learning disability	20	8
• Speech Impairment	40	17
• Mental Retardation	20	8
• Emotional Disturbance	<10	<4
• Hearing Impairment	20	8
• Visual Impairment	20	8
• Orthopedic Impairment	20	8
• Other Health Impairment	<10	<4
• Autism	30	13
• Traumatic Brain Injury	<10	<4
• Multiple Disabilities	40	17

• Deaf/Blindness	<10	<4
Grade		
• 7	<10	<5
• 8	30	17
• 9	30	17
• 10	40	22
• 11	30	17
• 12	20	11
• Ungraded	30	17

Note. According to the data-use agreement with IES, sample sizes are rounded to the nearest 10, and percentages are based on rounded numbers. The percentages listed are valid percentages, meaning they are based only on the participants that responded to each question.

Variables

Latent variables were created for transition planning, adaptive behavior, parent expectations, and the PSO. Observed variables came from parent survey Waves 1 - 5, school program survey Waves 1 and 2, and Table 2-3 presents the list of variables and their sources. The descriptions of the latent variables are presented below.

Table 2-3

List of NLTS2 Variables Used and Their Sources

Variable name	Description	NLTS2 variable name	Data sources
	Transition planning		
Transition goals	The variable indicated if the student had competitive employment, postsecondary education, and independent living as transition goals in school program surveys W1 or W2	npr1E4a, npr1E4c, npr1E4f, npr2E4a, npr2E4c, npr2E4f,	N2W1 School Program N2W2 School Program
Youth involvement	The variable indicated youth involvement as attending meetings and taking a leadership role in transition planning.	np1E3b, np1E2b, np1E2d	N2W1 Parent
	Adaptive behaviors		
Functional mental skills	Functional mental skill scale sumed up how well youth looks up telephone numbers, tells time, reads, and understands signs and counts change.	np1MentalSkill	N2W1 Parent
Self-care skills	The self-care ability scale was the sum of how well youth dresses and feeds themselves.	np1SelfCareSkill	N2W1 Parent

Social skills	Social skills scare sum of social assertation, self-control, and cooperation, and the scale was between 0-22	np1SocialSkills	N2W1 Parent
	Parent expectations		
Youth will attend postsecondary school	Parent expected youth will attend postsecondary school.	np1J2 (recoded)	N2W1 Parent
Youth will live away from home	Parent expected youth will live away from home.	np1J7 (recoded)	N2W1 Parent
Youth will get a job	Parent expected youth will get a job.	np1J9 (recoded)	N2W1 Parent
	PSO		
Postsecondary education	Student attended postsecondary education	np3D4a1_D4a2_D4a3_ever, np4D4a1_D4a2_D4a3_ever, np5A3a_A3e_A3i_ever	N2W3 Parent N2W4 Parent N2W5 Parent
Competitive employment	Students was competitively employed	W2CompEmplmt, np3CompEmplmt, np4CompEmplmt, np5CompEmplmt	N2W2 Parent N2W3 Parent N2W4 Parent N2W5 Parent
Independent living	Student lived away from home, including alone, with a spouse/roommate, or in a college dorm.	np5A1a_0506, np5A1a_0809	N2W5 Parent

Transition Planning

The transition planning construct was measured based on transition goals and youth involvement in transition planning as provided in NLTS2 surveys. Transition goals indicated that if youth has postsecondary education, competitive employment, and independent living goals in their transition planning, this scale was from 0 to 3. Transition goals were measured by school program survey Waves 1 and 2, and both data were combined to reduce missing values. On a scale of 0 to 3, a rating of 3 indicated students have goals of PSO, including postsecondary education, employment, and independent living. Youth involvement in transition planning was measured based on the following: youth attended IEP meetings and met teachers to set post-graduation goals and youth role in IEP planning, and it was on a 0-6 scale.

Adaptive Behaviors

The adaptive behaviors construct was conceptualized based on three variables, functional mental skills scale, self-care ability scale, and social skills as provided in NLTS2 surveys. The functional mental skill scale quantifies how well youth looks up telephone numbers, tells time, reads, and understands signs and counts change. The functional mental skills scale is between 4-16, where the score of 4-8 is considered low, 9-14 medium, and 15-16 high ability. The self-care ability scale was the sum of ratings on how well youth dresses and feeds themselves. The self-care ability scale is between 2-8, where the score of 2-4 is considered low; 5-7 medium, and 8 high abilities. The social skills scale measures social assertion, self-control, and cooperation, and the scale ranged 0-22. A score of 0-10 is the low ability, 11-16 is medium, and 17-22 is the high ability.

Parent Expectations

Parents rated the likelihood of their child attaining specific PSO in NLTS2 Wave 1. Parent expectations indicated parents' expectations regarding their child's postsecondary education, competitive employment, and independent living. The parent expectations construct included three items, and these questions were on a scale of 1-4 from "Definitely will" to "Definitely will not." These items were reverse coded so that higher scores present greater expectations as 1 indicated "*Definitely will not,*" and 4 indicated "*Definitely will*."

Postschool Outcome

PSO are defined in this study as postsecondary education, competitive employment, and independent living. Outcome variables data came from the NLTS2 parent/youth surveys (Wave 2 -5). Postsecondary education was measured as a binary variable of whether a student attends any type of postsecondary education, including vocational schools, two-years, or four years

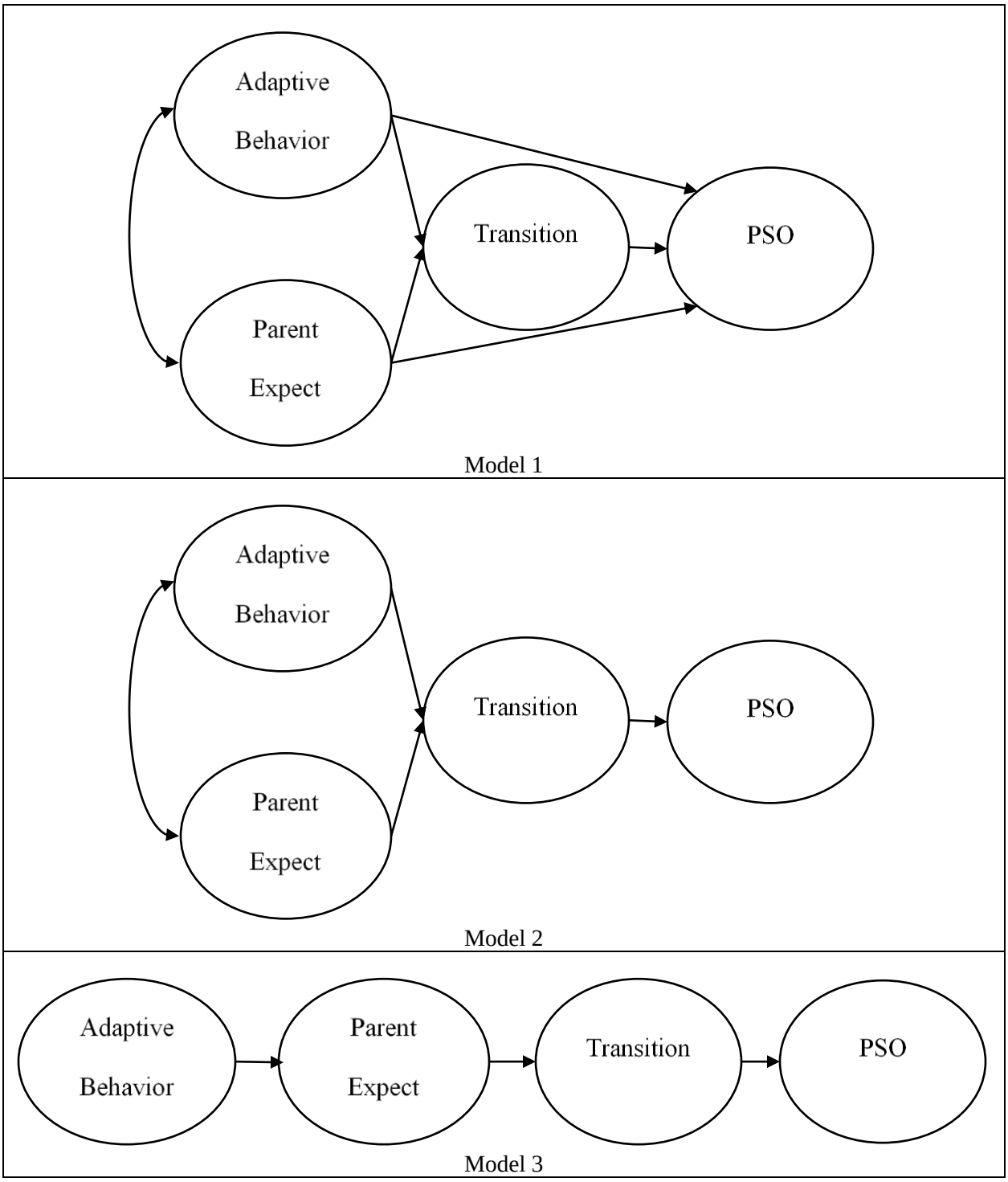
institutions since leaving high school. Variables were combined from parent/youth survey Waves 3 – 5 to reduce missing data on outcome variables. Competitive employment was identified based on two criteria: earning at least minimum wage and the majority of the workforce does not have a disability, according to the NLTS 2 study. Employment was also measured as a binary variable if a student was competitively employed or not in the last two years in parent/youth surveys. Competitive employment variables from parent/youth survey Wave 2-5 were combined to create employment variables with maximum valid cases. Independent living was defined as students' is living alone, with a spouse/roommate, or in a college dorm. Independent living data came from parent/youth survey Wave 5.

Statistical Analysis

Structural equation modeling (SEM) was used to test the relationships among adaptive behavior, parent expectations, transition, and PSO. SEM allows more complex modeling for the relationship among the multiple variables. Specifically, three models were tested (see Figure 2-1). In Model 1, adaptive behavior and parent expectations were hypothesized to directly and indirectly affect PSO. Transition served as the mediating latent variable. For Model 2, adaptive behavior and parent expectations did not directly affect PSO but had an indirect effect through the transition. Model 3 hypothesized that adaptive behavior positively affected parent expectations, parent expectations positively impacted transition, and transition positively impacted PSO. A two-step process was used to test the *a priori* models. The confirmatory factor analysis (CFA) was examined first to determine the quality of the measurement model. Next, the three structural models were tested.

Figure 2-1

The Three Models Tested



Data management and descriptive analyses were conducted using the Statistical Package for Social Sciences, Version 26. Missing values found to be completely random according to Little's Missing Completely at Random test ($\chi^2_{(180)} = 191.35$, $p = .34$); the degree of freedom was rounded to the nearest 10 based on IES data agreement. The missingness rates for the variables were between 4% to 26%. Missing data were imputed 20 times using Stata's ICE (imputation by chained equations) procedure (Royston, 2004, 2009; Royston et al., 2009).

SEM analysis was conducted with *R lavaan* package (Rosseel, 2012). The models were estimated using a weighted least squares estimator with a diagonal weight matrix (WLSMV) because this estimation is suggested for the models with categorical or ordinal variables (Brown, 2006; Muthen, 1984). Goodness-of-fit was assessed based on multiple fit indices that are χ^2, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Standardized Root Mean Squared Residual (SRMR). RMSEA values of less than .06, CFI values of more than .95, and SRMR values of less than .10 indicated a good fit (Hu & Bentler, 1999; Kline, 2016).

Results

Tests for Group Differences

Since 40% of our sample of 400 did not have any PSO data, we compared the group that was missing PSO data to those with PSO data on key variables and demographic variables to determine the equivalence of data with PSO and without PSO. We used the χ^2 test of independence for binary variables and *t*-test for interval and continuous variables to determine differences. There was no difference regarding gender, age, goal setting, youth involvement, functional mental health skills, selfcare skills, social skills, parent expectations of independent living, and getting a paid job. The results indicated that students who dropped from the survey

were more likely to be White, $\chi^2_{(1)} = 7.8$, $p = .005$, and their parents had a higher expectation of postsecondary education $t_{(290)} = 2.08$, $p = .039$. Those without PSO data came from lower-income households $t_{(190)} = -2.76$, $p = .006$ and their parents had lower levels of education $t_{(300)} = -3.15$, $p = .002$. As the data-use agreement with the Institute of Education Sciences requires, degrees of freedom were rounded to the nearest 10.

Table 2-4

Correlation Matrix of the Variables in the Model

	Transition planning			Adaptive behaviors		Parent expectations			PSO		
	1	2	3	4	5	6	7	8	9	10	11
1.Transition goal setting											
2.Youth involvement in planning	.23**										
3. Functional mental skills scale	.51**	.28**									
4. Self-Care skills scale	.25**	.15*	.48**								
5. Social skills scale score (sum all social scores)	.22**	.24**	.27**	.2**							
6.Youth will attend postsecondary school	.47**	.23**	.48**	.21**	.33**						
7. Youth will live away from home without supervision	.49**	.22**	.58**	.29**	.29**	.48**					
8. Youth will eventually get a paid job	.39**	.27**	.50**	.45**	.31**	.37**	.5**				
9. Competitively employed	.24**	.27**	.40**	.30**	.15*	.35**	.39**	.39**			
10. Postsecondary education	.32**	.08	.17**	.1	.16*	.32**	.17**	.25**	.16*		
11. Independent living	.31**	.13*	.19**	.17**	.04	.12	.26**	.18**	.22**	.14*	
Mean	0.97	2.77	10.86	7.20	12.92	2.52	2.38	3.38	0.47	0.39	0.15
SD	1.04	1.88	3.81	1.45	3.94	1.02	1.08	0.77	0.50	0.49	0.3

Descriptive Statistics of Observed Variables

Correlation coefficients, means, and standard deviations for the observed variables are reported in Table 2-4. All the correlation coefficients were positive, and the majority of them was

statistically significant. Correlation coefficients ranged between .04 and .58. The lowest correlation was between independent living and social skills, while the highest correlation coefficient was between self-care skills and parent expectations for independent living. Social skills had the highest mean of 12.92 and independent living had the lowest mean of 0.15.

Structural Equation Modeling Results

The goodness-of-fit statistics for the CFA and the three models are reported in Table 2-5. Results of the CFA suggested the measurement model was a reasonable fit data [$\chi^2_{(39)} = 86.35$, $p < .001$; RMSEA = .072 (90% CI: 051, .092); CFI = .93; SRMR = .066], and there were high correlations among the latent variables, ranging from .75 (adaptive behavior with PSO) to .95 (adaptive behavior with parent expectations). The regression weights between the observed variables and the latent variables were all statistically significant. For model 1, [$\chi^2_{(39)} = 91.01$, $p < .001$; RMSEA = .049 (90% CI: 022, .072); CFI = .98; SRMR = .06], the global fit appeared reasonable, but none of the structural paths were statistically significant. It was noted that the standard errors for structural coefficients were very large, which results in nonsignificant structural coefficients. Similar results were found for model 2, with a reasonable global fit [$\chi^2_{(41)} = 92.49$, $p < .001$; RMSEA = .046 (90% CI: 018, .069); CFI = .98; SRMR = .06], but it included a nonsignificant structural path. Model 3 was the only model with a reasonable fit [$\chi^2_{(42)} = 93.52$, $p < .001$; RMSEA = .044 (90% CI: 016, .067); CFI = .98; SRMR = .06] and all structural paths were statistically significant.

Table 2-5

The Goodness of Fit Statistics for All Tested Measurement Models

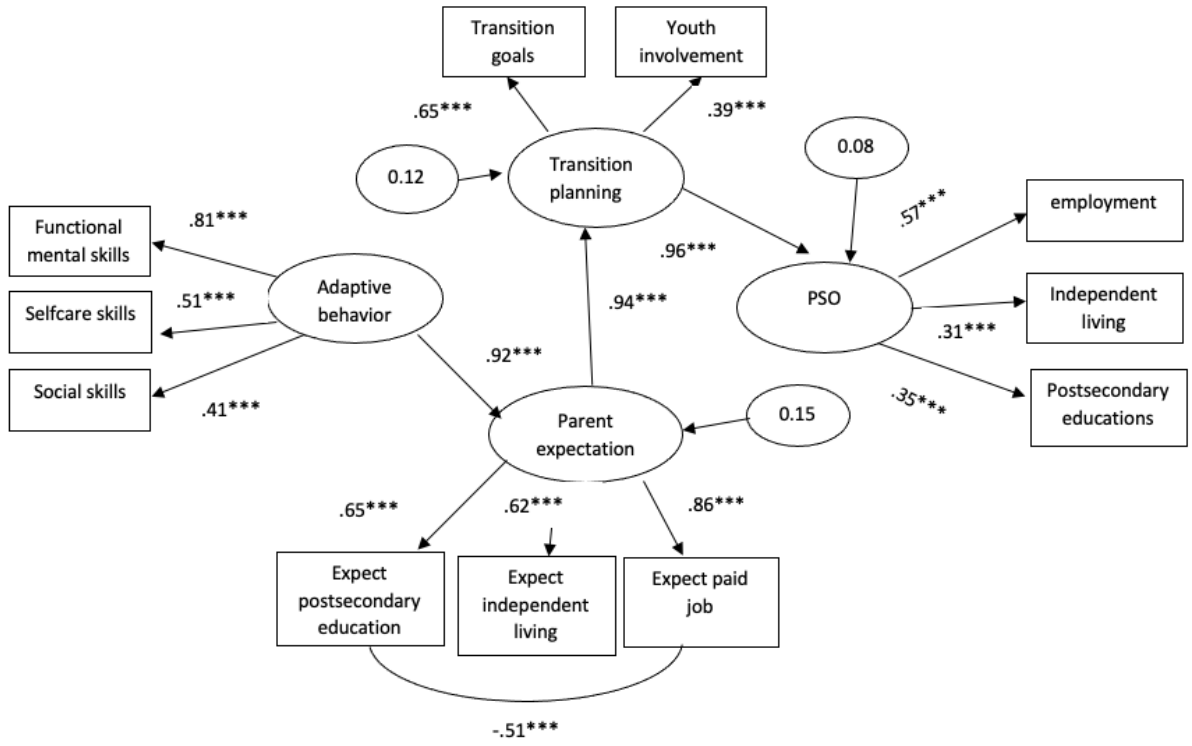
Model	χ^2	*df*	RMSEA	[90% CI]	CFI	SRMR
CFA	86.35	39	.072	[.051, .091]	.93	.066
Model 1	91.01	39	.049	[.022, .072]	.98	.060
Model 2	92.49	41	.046	[.018, .069]	.98	.060
Model 3	93.52	42	.044	[.016, .067]	.98	.060

Note. All models are statistically significant at the $p < .001$ level. RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual.

Figure 2-2 provides the standardized parameter estimations for model 3. The direct and indirect unstandardized and standardized regression coefficients are reported in Table 2-6. All the indirect effects were statistically significant. The indirect effects of adaptive behavior on POS (adaptive behavior → parental expectations → transition → PSO) indicted that one standard deviation increase for the latent variable of adaptive behavior resulted in a .82 standard deviation increase in PSO. Parental expectations had an indirect effect on PSO (parental expectations → transition → PSO), with one standard deviation increase in parental expectations resulting in a .90 standard deviation increase in PSO. The transition directly affected PSO, with one standard deviation increase in transition resulting in a .96 increase in PSO. Model 3 explained 85% of the variance in parent expectations, 88% of the variance in transition planning, and 92% of PSO variance.

Figure 2-2

Model 3: Standardized Estimates are Listed for Displayed Model Paths



Note. RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual; PSO = Postschool outcomes. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. Model fit: $\chi^2_{(42)} = 93.52$, $p < .001$; RMSEA = .044 (90% CI: .16 – .067); CFI = .985; SRMR = .06.

Table 2-6

Unstandardized Estimates and Standard Errors of Direct and Indirect Effects for Model 3

	Estimate	SE	Std. Estimates
Direct effect			
Adaptive behavior -> Parent expectations	.83	.12	.92
Parent expectations-> Transition planning	.96	.08	.94
Transition planning -> PSO	.41	.05	.96
Indirect effect			
Adaptive behavior -> PSO	.32	.04	.82
Adaptive behavior -> Transition planning	.79	.13	.86
Parent expectations-> PSO	.39	.05	.90

Note. All estimations are statistically significant at the $p < .001$ level. Std. = Standardized, SE = Standard error, PSO = Postschool outcomes.

Discussion

This study examined the underlining relationship between adaptive behavior, parent expectations, transition planning, and PSO for EL students with disabilities. Although the majority of previous research examined PSO with regression analysis, some recent research, including our study, examined PSO using SEM modeling for SWD (e.g., Dell'Armo & Tassé, 2019; Kirby, 2016; Shogren et al., 2018). The present study included testing the use of four latent variables (i.e., adaptive behavior, parent expectations, transition planning, and PSO) to fit the model predicting EL with disabilities' PSO. The results indicated that adaptive behavior is a significant predictor of parent expectations, and there was a significant indirect effect of adaptive behavior on transition planning and PSO. Similarly, parent expectations significantly directly affected transition planning and a significant indirect effect on EL with disabilities' PSO. This study expanded the literature regarding EL with disabilities' PSO and the importance of parent expectations and transition planning.

Adaptive behaviors were significant predictors of parent expectations (Kirby, 2016; Dell'Armo &Tassé, 2019). This study demonstrated similar results with previous research regarding those adaptive behaviors (self-care skills, functional mental skills, and social skills) were significant predictors of parent expectations (parent expectations for postsecondary education, employment, and independent living) for EL with disabilities. Kirby (2016) reported that adaptive behavior has a significant indirect effect on PSO among autistic students, and Dell'Armo and Tassé (2019) presented the direct effect of adaptive behaviors on PSO. Our study expanded the literature regarding the indirect effect between adaptive behaviors and transition planning. The results suggested that adaptive behavior had an indirect effect on PSO. Our study provided empirical evidence that adaptive behaviors, which does not have a direct effect on

transition planning and PSO, had a significant indirect effect on transition planning (transition goals and youth involvement) and PSO.

Extensive research has shown that parent expectations are a significant predictor of PSO (Carter et al., 2012; Kirby, 2016; Trainor et al., 2019; Wagner et al., 2014). In contrast, some research indicated that parent expectations were not a significant predictor of employment (Chiang et al., 2013) and PSO (Dell'Armo & Tassé, 2019) for autistic students and intellectual disability students. Our study showed an indirect relationship between parent expectations and PSO through transition planning. This result could explain the insignificant direct effect of parent expectations to PSO but indicates that effect is mediated by transition planning.

Previous research has established that transition planning is crucial for a successful postschool transition (Mazzotti et al., 2016; Mazzotti et al., 2021; Trainor et al., 2019). This study presented that transition planning significantly predicted EL with disabilities' PSO (postsecondary education, employment, and independent living) and was a mediator for the effects of adaptive behavior and parental expectations on PSO. It was known that almost all EL with disabilities had transition planning and their transition was aligned with courses even though postschool employment was lower for EL with disabilities (Trainor et al., 2016). Our results also suggested a complex relationship between transition goals, youth involvement, and PSO for EL with disabilities.

Limitations and Future Directions

Although the results of the study were promising, there were some limitations. First, this study was designed as a secondary analysis of NLTS2. Thus, the study was limited to variables and scales (binary, ordinal or continuous) as provided in the dataset. For instance, PSO, postsecondary education, and employment were measured on the binary scale as yes or no.

Similarly, Trainor et al. (2020) suggested postsecondary education and employment may not be sufficient to measure postschool transition and they recommended focusing on the quality of life as an indicator of a successful postschool transition. Second, EL with disabilities sample in the NLTS2 dataset was small, and a large percentage of the sample (40%) was lost because they did not have PSO, which limited the generalizable of the results. Third, NLTS2 data used a two-stratified sampling and provided replication weights. We did not used weights in this study because the research purpose was to examine the underlining relationship between PSO, transition planning, adaptive behavior, and parent expectations. Researchers may use replication weights with R *lavaan survey* package to examine SWD's or EL with disabilities PSO to eliminate a potential sampling bias.

Future research should validate this model once NLTS2012 postschool data become available. Although EL with disabilities were identified based on transcript and school program surveys in this study, EL with disabilities were identified by the school districts in NLTS2012. Our results demonstrated direct and indirect effects among adaptive behavior, parent expectations, transition planning on postschool outcomes. Future studies may examine the internal relationship among these variables as Trainor et al. (2020) indicated that transition is a complex and multilayer process. For example, examining the change in parent expectations after SWD receive an intervention on improving students' adaptive behaviors may reveal the relationship between parent expectation and adaptive behaviors. In addition, assessing the parent expectation at the beginning of the high school and at the end of high school could present the influence of transition programs on parent expectations.

Disability types such as high and low incidences may be related to EL with disabilities' PSO. Thus, research should explore EL with disabilities' postschool transition based on their

disability categories. As Dell'Armo and Tassé (2019) suggested, this model can be used to analyze other disability groups and their PSO. The research on EL with disabilities and their postschool transition should investigate evidence-based intervention for a successful postschool transition.

Implications

This research examined the relationship between adaptive behaviors, parent expectations, transition planning, and PSO for EL with the disabilities' student population. These results have implications for teachers, students and their parents. An important implication is that recognizing supporting students' adaptive behavior could lead better postschool outcomes through parent expectations and transition planning. In this sense, this study reiterated the importance of collaboration among students, parents, and teachers. This means understanding the students' need should be a start point for transition planning and the desired PSO. Parents and teachers should focus on improving students' adaptive behavior.

Considering adaptive behaviors significantly predicted parent expectations, transition planning, and PSO, this study highlighted the significance of adaptive behaviors. Thus, interventions on improving adaptive behaviors could support EL with disabilities parent expectations, transition planning, and PSO. Our study showed that researchers need to consider the intersectional relationships among these variables when conducting research and implementing interventions. Besides, the results highlighted the importance of parent expectations and transition planning for EL with disabilities. The stakeholders (parents, policymakers, students, and teachers) should invest time and resources in parent expectations, primary transition goals of postsecondary education, employment, independent living, and youth

involvement in the transition program to support a successful postschool transition EL with disabilities.

References

Artiles, A. J., & Klingner, J. K. (2006). Forging a knowledge base on English language learners with special needs: Theoretical, population, and technical issues. Teachers College Record, 108(11), 2187.

Baker, L. L. (2017). What It Takes to Succeed: The Importance of Social Support for Academically Successful Middle School English Learners. Youth & Society, 49(5), 658–678.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Calderón, M., Slavin, R. E., & Sánchez, M. (2011). Effective instruction for English learners. *Future of Children, 21*, 103–127.

Cameto, R., Levine, P., and Wagner, M. (2004). Transition Planning for Students With Disabilities. Menlo Park, CA: SRI International.

Carter, E. W., Austin, D., & Trainor, A. A. (2012). Predictors of postschool employment outcomes for young adults with severe disabilities. *Journal of Disability Policy Studies, 23*(1), 50–63. https://doi.org/10.1177/1044207311414680.

Center for Change in Transition Services. (2021). Post-school outcomes. https://www.seattleu.edu/ccts/post-school-outcomes/

Chiang, H.-M., Cheung, Y. K., Li, H., & Tsai, L. Y. (2013). Factors associated with participation in employment for high school leavers with autism. *Journal of Autism and Developmental Disorders*, 43, 1832–1842.

Daviso, A. W., Denney, S. C., Baer, R. M., & Flexer, R. (2011). Postschool goals and transition services for students with learning disabilities. American Secondary Education, 77-93.

Dell'Armo, K. A., & Tassé, M. J. (2019). The role of adaptive behavior and parent expectations in predicting post-school outcomes for young adults with intellectual disability. *Journal of autism and developmental disorders*, *49*(4), 1638-1651. https://doi.org/10.1007/s10803-018-3857-6

Flores, S. M., Park, T. J., Viano, S. L., & Coca, V. M. (2017). State policy and the educational outcomes of English learner and immigrant students: Three administrative data stories. *American Behavioral Scientist, 61*, 1824–1844. doi:10.1177/0002764217744836

Greene, G. (2014). Transition of culturally and linguistically diverse youth with disabilities: Challenges and opportunities. *Journal of Vocational Rehabilitation, 40*(3), 239–245. doi:10.3233/JVR-140689

Hoover, J. J., & Patton, J. R. (2005). Differentiating curriculum and instruction for English-language learners with special needs. *Intervention in School and Clinic, 40*(4), 231-235.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. HTTPS:// doi.org/10.1080/10705519909540118.

Individuals With Disabilities Education Act (IDEA) of 2004, 20, U.S.C. § 1400 (2004).

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record, 117*(12), 1-44.

Kirby, A. V. (2016). Parent expectations mediate outcomes for young adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(5), 1643–1655. https://doi. org/10.1007/s10803-015-2691-3.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford publications. (Fourth edition).

Kohler, P. D. (1996). *Taxonomy for transition programming: Linking research and practice.* Champaign, IL: University of Illinois.

Lipscomb, S., Haimson, J., Liu, A.Y., Burghardt, J., Johnson, D.R., & Thurlow, M.L. (2017). *Preparing for life after high school: The characteristics and experiences of youth in special education. Findings from the National Longitudinal Transition Study 2012. Volume 1: Comparisons with other youth: Full report (NCEE 2017-4016)*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Marcotte, D. E., Bailey, T., Borkoski, C., & Kienzel, G. S. (2005). The returns of a community college education: Evidence from the National Education Longitudinal Survey. *Educational Evaluation and Policy Analysis, 27*(2), 157–175. https://doi. org/10.3102/01623737027002157.

Mazzotti, V. L., Rowe, D. A., Kwiatek, S., Voggt, A., Chang, W. H., Fowler, C. H., Poppen, M. Sinclair, J., & Test, D. W. (2021). Secondary Transition Predictors of Postschool Success: An Update to the Research Base. *Career Development and Transition for Exceptional Individuals, 44*(1), 47-64. https://doi.org/10.1177/2165143420959793

Mazzotti, V. L., Rowe, D. A., Sinclair, J., Poppen, M., Woods, W. E., & Shearer, M. L. (2016). Predictors of post-school success: A systematic review of NLTS2 secondary analyses. *Career Development and Transition for Exceptional Individuals, 39*(4), 196-215. https://doi. org/10.1177/2165143415588047

McDowell, L. (2004). Work, workfare, work/life balance and an ethic of care. *Progress in Human Geography*, *28*(2), 145-163. https://doi. org/10.1191/0309132504ph478oa

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., and Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017- 144). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144.

Murphy, A. F., & Torff, B. (2019). Teachers' Beliefs about Rigor of Curriculum for English Language Learners. *Educational Forum, 83*(1), 90–101.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132. https://doi.org/10.1007/BF022 94210.

Newman, L. A., & Madaus, J. W. (2015). An analysis of factors related to receipt of accommodations and services by postsecondary students with disabilities. *Remedial and Special Education, 36*(4), 208–219. https://doi. org/10.1177/0741932515572912

Newman, L. A., Madaus, J. W., & Javitz, H. S. (2016). Effect of transition planning on postsecondary support receipt by students with disabilities. Exceptional Children, 82(4), 497-514.

Newman, L., Wagner, M., Cameto, R., & Knokey, A. M. (2009). The Post-High School Outcomes of Youth With Disabilities up to 4 Years After High School: A Report From the National Longitudinal Transition Study-2 (NLTS2). NCSER 2009-3017. National Center for Special Education Research.

Newman, L., Wagner, M., Knokey, A., Marder, C., Nagle, K., Shaver, D., & Wei, X. (2011). *The post-high school outcomes of young adults with disabilities up to 8 years after high*

school: A report from the national longitudinal transition study-2 (NLTS2) (NCSER 2011-3005). Menlo Park, CA: SRI International.

Papay, C. K., & Bambara, L. M. (2014). Best practices in transition to adult life for youth with intellectual disabilities. *Career Development for Exceptional Individuals, 37*(3), 136–148. https://doi. org/10.1177/2165143413486693.

Park, J., & Bouck, E. (2018). In-school service predictors of employment for individuals with intellectual disability. *Research in Developmental Disabilities*, *77*(June2018), 68– 75. https://doi.org/10.1016/j.ridd.2018.03.014

Powers, K. M., Gil-Kashiwabara, E., Geenen, S. J., Powers, L. E., Balandran, J., & Palmer, C. (2005). Mandates and effective transition planning practices are reflected in IEPs. Career Development for Exceptional Individuals, 28(1), 47-59. Chicago

Roessingh, H., & Douglas, S. (2012). English language learners' transitional needs from high school to university: An exploratory study. *Journal of International Migration and Integration*, 13(3), 285-301.

Rojewski, J. W., Lee, I. H., & Gregg, N. (2014). Intermediate work outcomes for adolescents with high-incidence disabilities. *Career Development and Transition for Exceptional Individuals*, *37*(2), 106–118. https://doi.org/10.1177/216 5143412473352

Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software*, 48(2), 1–36. https://www.jstatsoft.org/v48/i02/.

Rowe, D. A., Alverson, C. Y., Unruh, D., Fowler, C., Kellems, R., & Test, D. W. (2015). A Delphi study to operationalize evidence-based predictors in secondary transition. *Career Development and Transition for Exceptional Individuals*. Advance online publication. doi:10.1177/2165143414526429

Rowe, D. A., Mazzotti, V. L., Fowler, C. H., Test, D. W., Mitchell, V. J., Clark, K. A., Holzberg, D., Owens, T., Rusher, D., Seaman-Tullis, R. L., Gushanas, C. M., Castle, H., Chang, W., Voggt, A., Kwiatek, S., & Dean, C. (2021). Updating the secondary transition research base: Evidence-and research-based practices in functional skills. *Career Development and Transition for Exceptional Individuals*, *44*(1), 28-46. https://doi.org/10.1177%2F2165143420958674

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, *4*, 227–241.

Royston, P. (2009). Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. *Stata Journal*, *9*, 466–477.

Royston, P., Carlin, J. B., & White, I. R. (2009). Multiple imputation of missing values: New features for mim. *Stata Journal*, *9*, 252–264.

Schlaman, H. (2019). Designing structures and pathways to support language development and content learning for English Learners: Dilemmas facing school leaders. *International Multilingual Research Journal, 13*(1), 32-50.

Shattuck, P. T., Narendorf, S. C., Cooper, B., Sterzing, P. R., Wagner, M., & Lounds, J. (2012). Postsecondary education and employment among youth with an autism spectrum disorder. *Pediatrics*, *129*(6), 1042–1049. https://doi.org/10.1542/ peds.2011-2864

Shogren, K. A., Wehmeyer, M. L., Shaw, L. A., Grigal, M., Hart, D., Smith, F. A., & Khamsi, S. (2018). Predictors of self-determination in postsecondary education for students with intellectual and developmental disabilities. *Education and Training in Autism and Developmental Disabilities*, *53*(2), 146-159. https://www.jstor.org/stable/26495266

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick- Duffy, S. A., Spreat, S., Thissen, D., Widaman, K. F., & Zhang, D. (2012). The construct of adaptive behavior: Its

conceptualization, measurement, and use in the field of intellectual disability. *American Journal on Intellectual and Developmental Disabilities, 4*, 291–303. https://doi.org/10.1352/1944-7558-117.4.291.

Test, D. W., Fowler, C. H., Richter, S. M., White, J., Mazzotti, V., Walker, A. R., Kohler, P., & Kortering, L. (2009). Evidence-based practices in secondary transition. *Career Development for Exceptional Individuals, 32*, 115–128. doi:10.1177/0885728809336859

Test, D. W., Mazzotti, V. L., Mustian, A. L., Fowler, C. H., Kortering, L., & Kohler, P. (2009). Evidence-based secondary transition predictors for improving postschool outcomes for students with disabilities. *Career Development for Exceptional Individuals, 32*(3), 160-181.

Trainor, A. A., Carter, E. W., Karpur, A., Martin, J. E., Mazzotti, V. L., Morningstar, M. E., ... & Rojewski, J. W. (2020). A framework for research in transition: Identifying important areas and intersections for future study. *Career Development and Transition for Exceptional Individuals, 43*(1), 5-17. https://doi.org/10.1177%2F2165143419864551

Trainor, A. A., Murray, A., & Kim, H. (2014). *Postsecondary Transition and English Learners with Disabilities: Data from the Second National Longitudinal Transition Study.* WCER Working Paper No. 2014-4. Wisconsin Center for Education Research.

Trainor, A. A., Newman, L., Garcia, E., Woodley, H. H., Traxler, R. E., & Deschene, D. N. (2019). Postsecondary Education-Focused Transition Planning Experiences of English Learners with Disabilities. Career Development and Transition for Exceptional Individuals, 42(1), 43–55.

Trainor, A., Murray, A., & Kim, H. J. (2016). English learners with disabilities in high school: Population characteristics, transition programs, and postschool outcomes. *Remedial and Special Education, 37*(3), 146-158.

US Department of Justice & US Department of Education. (2015, January). *English learner (EL) dear colleague*. Retrieved from https://www2.ed.gov/about/offices/list/ocr/letters/colleague-el-201501.pdf

Wagner, M. M., Newman, L. A., & Javitz, H. S. (2014). The influence of family socioeconomic status on the post-high school outcomes of youth with disabilities. *Career Development and Transition for Exceptional Individuals, 37*(1), 5–17. https://doi. org/10.1177/2165143414523980.

Wanzek, J., Swanson, E., Vaughn, S., Roberts, G., & Fall, A. M. (2016). English learner and non-English learner students with disabilities: Content acquisition and comprehension. *Exceptional Children, 82*(4), 428–442. Retrieved from http://dx.doi.org/10.1177/0014402915619419

Watkins, K. (2015). *English language learners and high school transition experiences.* [Unpublished doctoral dissertation]. Walden University

Wei, X., Yu, J. W., Shattuck, P., & Blackorby, J. (2017). High school math and science preparation and postsecondary stem participation for students with an autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, *32*(2), 83–92. https://doi.org/10.1177/1088 357615588489

Wei, X., Wagner, M., Hudson, L., Yu, J. W., & Javitz, H. (2016). The effect of transition planning participation and goal-setting on college enrollment among youth with autism spectrum disorders. *Remedial and Special Education*, *37*(1), 3-14.

CHAPTER 3 [ARTICLE III]: APPLYING LATENT VARIABLE APPROACH FOR EXAMINING MEASUREMENT AND PREDICTION INVARIANCE IN ENGLISH LEARNERS LARGE-SCALE ASSESSMENTS

Gezer, T., Flowers, C., & Kim, S. Applying latent variable approach for examining measurement and prediction invariance in English learners large-scale assessments. *Journal of Educational Measurement*

Reducing the achievement gap among student groups has been the focus of educational policies in the United States since the enactment of No Child Left Behind (NCLB, 2001) and the Every Student Succeeds Act (ESSA, 2015). These federal policies required all students to participate in standardized testing regardless and interventions for closing the achievement gap among student groups implemented (Bunch, 2011; Heise, 2017). Student characteristics including gender, race/ethnicity, English learners (EL), students with disabilities (SWD), and economically disadvantaged students have been examined in achievement gap literature (Lee, 2002; Polat et al., 2016; Thurlow et al., 2016). According to American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME; 2014), reporting aggregated scores for subgroups (race, gender, EL, & SWD) requires test users to provide comparability evidence and add cautionary statements if necessary; they further suggested that potential biases in measurement and prediction for student subgroups should be investigated to ensure fairness in interpreting test scores.

Measurement and Prediction Invariance Models

Measurement invariance (MI) investigates the consistency of measures for different conditions such as time, population, or method (Meade & Lautenschlager, 2004). Examining the MI over populations focuses on construct bias among groups. Putnick and Bornstein (2016)

emphasized that MI is critical in psychological research because it requires comparing group means. Thus, examining MI is considered a standard procedure of large-scale test developments (AERA, APA, & NCME, 2014) to maintain a common interpretation and inference of assessment scores across diverse student groups.

Prediction invariance (PI) maintains the uniform interpretation of test results regardless of demographic characteristics such as gender, race, language status (AERA, APA, & NCME, 2014). A review on the PI research suggested that using observed test scores to predict future events may not be invariant for subgroups. For example, SAT and high school GPA are often used to predict first-year college performance but studies suggest the results may lead to overprediction of certain ethnic groups such as Black, Hispanic (Aguinis et al., 2016; Berry & Zhao, 2015; Mattern & Patterson, 2013) and underprediction of females' college grades compare to males (Fischer et al., 2013; Keiser et al., 2016). Considering that a typical aim of standardized tests is to predict academic or job performance (Cleary, 1968; Humphreys, 1952), PI becomes crucial because overprediction leads to unfair benefits, and underprediction causes unfair penalties (Culpepper et al., 2019).

Millsap (2007) stated that when tests are used for selection, prediction tends to gain more attention than measurement, which inadvertently leads to overlooking measurement concerns in the testing process. Psychometric research suggests that latent level differences could explain observed score group differences (Bryant, 2004; Millsap, 2007) because the latent variable model may provide an understanding of the unobserved process. The observed score is defined as the sum of true score and error according to classical test theory (Lord & Novick, 1968, p. 30). On the other hand, latent variable models can control measurement error better than observed variable models (Kline, 2016). Culpepper et al. (2019) stated that observed scores provide a

limited representation of what happens under the surface (p.300) and suggested that latent variable models can be used to analyze MI, PI, and the latent structure for subtests. Thus, this study used both latent scores and observed scores to analyze EL large-scale assessment data.

It is important to examine MI and PI for EL students. Jonson et al. (2019) analyzed 18 academic and intelligence tests based on AERA, APA & NCME (2014) fairness standards. Although MI and PI are necessary aspects of test fairness, they concluded that the EL student group had been overlooked in both preliminary and standardization sampling in examining test fairness process. Therefore, this study aims to examine MI and PI among EL students using large-scale assessment data.

Literature Review

English Learners in the US Schools

The number of EL students continues to grow in the US. While about 10% of K-12 students are EL, the proportion of EL is expected to reach 25% by 2025 (National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs, 2007). Although most EL is born in the US, some EL immigrated to the US (Flores et al., 2017). Thus, immigration may have interrupted some EL students' academic progress (Calderón et al., 2011). Even though state accountability systems tended to categorize the EL population as binary, EL and non-EL, these student populations demonstrated a great amount of diversity in terms of language proficiency in English, their native language, and their academic proficiencies (Abedi & Linquanti, 2012). Considering the diverse characteristics of the EL population, it is necessary to address this heterogeneity in EL education (National Academies of Sciences, Engineering, and Medicine, 2017, 2018). Instead of comparing EL and non-EL students' academic achievement, this study takes a closer look at the measurement and prediction properties of an EL large-scale

assessment in terms of ethnicity and time spent in US schools. While many studies examined the achievement gap between EL and non-EL students in large-stake assessments, such as National Assessment of Educational Progress and state accountability assessments in reading and math (Abedi, 2006; Fry, 2007; Miley & Farmer, 2017; Polat et al., 2016), less research has been conducted examining the heterogeneity within the EL population.

English Language Proficiency and English Language Art Assessments

English language proficiency (ELP) is necessary for EL's academic success in the US educational system. Corcoran (2018) highlighted that EL students struggle with academic performance if they are not proficient in English. Following instruction, contributing to class discussions, expressing ideas in writing, or demonstrating knowledge on testing requires ELP, and EL students may struggle to complete these activities due to limited English proficiency (Clausen, 2017; McArdle et al., 2019).

EL students are required to take state accountability assessments in reading and math (Wright, 2006). EL students take English language Art (ELA) assessments from 3rd to 8th grade and once in high school. Parker et al. (2009) analyzed the relationship between English language proficiency (ELP) domains (i.e., reading, writing, speaking, and listening), and content assessments (i.e., reading and math), by using data from three states. They concluded that English language domains explained 14%-30% of the variance of content scores, and they further indicated that reading and writing domains of ELP were better predictors of reading achievement than listening and speaking domains.

He (2021) examined the relationship between English proficiency and ELA achievement among EL students in a school district in New York. The results suggested that English proficiency was a significant predictor of ELA achievement, and gender and a grade level did not

influence the relationship between English proficiency and ELA achievement. The author recommended integrating educational activities to support EL students' language and literacy development.

Factors Associated with EL's ELA Achievement

English Language Proficiency

According to Cummins (1984), there are two types of language proficiency in a second language: basic interpersonal communication skills and cognitive-academic language proficiency. In this study, ELP indicates cognitive-academic language proficiency in English. According to the Every Student Succeeds Act (ESSA, 2015), states need to implement ELA standards that require measuring language proficiency in four domains (i.e., speaking, reading, writing, and listening) and aligning language proficiency with state academic standards (US Department of Education, 2015, p. 24). EL students may reach English language speaking and listening proficiencies within two years, but it can take up to seven years to attain English writing and reading proficiencies (Thompson, 2017).

World-Class Instructional Design and Assessment (WIDA) was created as a response to NCLB to assess ELP. WIDA is the common ELP assessment as 41 US states, territories, and federal education have adopted WIDA English language development standards (WIDA Consortium, 2021, July 26). The WIDA ACCESS aims to measure the developing English language proficiency of EL in Grades K-12 in the United States (WIDA Consortium, 2012). ACCESS assesses social, instructional, and academic English in ELA, math, science, and social studies (WIDA Consortium, 2020). The same report indicated that WIDA test scores can be used for accountability, reclassifying students for exiting English language support services, deciding the instructional planning, and monitoring students' English language progress.

Hispanic EL Students

The number of Hispanic EL students is increasing (Kena et al., 2016). Considering the specific characteristics of Hispanic families may explain the potential academic risks that EL Hispanic students may encounter. For instance, the majority of the Hispanic population (73%) speaks Spanish at home (Krogstad, 2016), and 85% of Hispanic parents indicated talking with their kids in Spanish (Lopez et al., 2018). These statistics are important because they may influence Hispanic students' ELP. About 76% of Hispanic students graduated from high school in four years, and this percentage is a little higher than Black students' graduation rate (68%) but lower than White (85%) and Asian/Pacific Islander students (93%; National Center for Education Statistics, 2014). Also, Hispanic EL students are more likely to come from economically disadvantaged households (Fry & Gonzalez, 2008). The economic disadvantages could influence Hispanic students' academic performance, which in turn possibly leads to the lower high school graduation rate. After controlling for all other factors, Spanish speakers and students with lower parent education had a lower probability of being reclassified as English proficient (Slama, 2014; Thompson, 2017). It is necessary to examine potential biases in the measurement process reflecting the characteristics of Hispanic households. Thus, looking at this from psychometric research perspectives, lack of measurement invariance could be a reason for the slow progression to ELP.

Years in the US Schools

An important indicator of ELP has been years in the US (Portes & Hao, 1998), as students would be exposed to the English language for this time (Jia & Aaronson, 2003). However, Slama (2012) inserted that many US-born EL students started high school without reaching ELP, which means they spent nine years as EL. The same study examined the

longitudinal trajectories of English academic proficiency for high school students in a northeastern state in the US. The results indicated that US-born EL students had higher ELP than foreign-born EL students at the beginning of high school. Although foreign-born EL students reach a similar level of ELP to US-born EL students by the end of high school, both groups have lower academic English proficiency throughout high school, which means both groups had limited educational opportunities.

Another aspect of years in the US schools is long-term English learners (LTEL), and they are defined as EL students who were not reclassified as English proficiency within five years (ESSA, 2015). LTEL usually have medium or low ELP (Slama, 2014) and low academic achievement (Callahan, 2005; Flores et al., 2012). Some argue that external factors such as inadequate services and bureaucratic errors cause LTEL rather than their ELP (Brooks, 2018; Flores et al., 2015). In any case, examining MI and PI across the years in the US schools may provide some insights regarding LTEL.

Research Questions

This study analyzes MI and PI simultaneously based on latent scores and compares the results with observed scores prediction using EL large-scale testing data. The following research questions have guided this study:

1. Are ELP measures (i.e., listening, speaking, reading, and writing) invariant across Hispanic vs. non-Hispanic EL students?
2. Are ELP measures of prediction of ELA achievement invariant across Hispanic vs. non-Hispanic EL students?
3. Are ELP measures (i.e., listening, speaking, reading, and writing) invariant across years in the US schools (three years or less and more than three years)?

4. Are ELP measures of prediction of ELA achievement is invariant across years in the US schools?
5. Is there a difference between a latent score and observed score prediction of ELA achievement?

Method

Study Design and Sample

Data for this study came from a southeastern state 2019 ELA accountability assessment and WIDA's ACCESS ELP assessment. The data were de-identified in three separate files, (a) demographic data, (b) WIDA ACCESS assessment results, and (c) ELA accountability assessment results. These files were merged based on the provided Encrypted Student Identification number. Only 3rd, 8th, and 10th grades data were used in this study. We selected a grade from each school level, elementary, middle and high school based on adequate sample size to conduct the statistical analysis.

The demographic characteristics of the students are reported by a grade level in Table 3-1. The majority of the sample across all grades was male (54%-58%), Hispanic (77%-80%), economically disadvantaged (65%-68%), and non-immigrant students (98%-99%). For three years or less, the percentages of students in US schools were 14.7%, 62.1%, and 61.7% for 3rd, 8th and 10th grades, respectively.

Table 3-1

Demographic Characteristics of Sample Based on Grade Levels

	3rd (*N*=9718)		8th (*N*=4981)		10th (*N*=3139)	
	F	%	F	%	F	%
Gender						
Female	4427	45.5	2100	42.2	1402	44.7
Male	5291	54.3	2881	57.8	1734	55.2
Economically Disadvantaged						

No	3376	34.7	1635	32.8	1003	32
Yes	6342	65.3	3346	67.2	2133	68
Migrant						
No	9605	98.8	4902	98.4	3110	99.2
Yes	113	1.2	79	1.6	26	0.8
Ethnicity						
Asian	1081	11.1	410	8.2	301	9.6
Black	270	2.8	255	5.1	210	6.7
Hispanic	7794	80.2	3993	80.2	2436	77.6
American Indian	15	0.2	11	0.2	4	0.1
Multiracial	59	0.6	39	0.8	20	0.6
Pacific Islander	22	0.2	18	0.4	11	0.4
White	477	4.9	255	5.1	154	4.9
Years in the US Schools						
≤ 3 years	1431	14.7	3095	62.1	1937	61.7
> 3 years	8307	85.5	1898	38.1	1204	38.4

Measures

ELA Assessment

ELA tests are administered annually to all students in public schools. ELA assessments are multiple-choice exams and are aligned with the North Carolina Standard Course of Study. The testing window for this assessment is the last 10 days of instructional days of the school year. While all students have to take the ELA assessment, EL students could be exempt from ELA assessment if they have been enrolled in US schools less than 12 months. Students who did not take the ELA assessment were not included in the final sample. ELA scale scores range between 406-461; 422-479; and 118-167 for 3rd, 8th and 10th grades, respectively (see Table 3-2).

English Language Proficiency Assessment

Every Student Succeeds Act (ESSA, 2015) requires states to assess ELP in four domains: listening, reading, speaking, and writing. A student's language proficiency is initially determined by the WIDA Screener assessment and subsequently measured annually by the WIDA ACCESS

until a certain level of language proficiency is attained. WIDA ACCESS results are used to estimate students' English academic language and facilitate English language development support.

WIDA-ACCESS is a computer adaptive test and is given annually to EL students. Item-level data was not available and, thus, domain-level information was used in this study. Grant and Kraninger (2019) suggested that reporting individual language domain scores and an overall composite score is adequate for WIDA-ACCESS (As cited in WIDA, 2020 p.4-2). Thus, domain-level data was used to test MI. The WIDA-ACCESS assessment reports scale scores and proficiency levels for each language domain. The scale scores are between 100 and 600, and the proficiency levels range from 1 to 6. The proficiency levels from 1 to 6 are labeled as entering, emerging, developing, expanding, bridging, and reaching. Proficiency levels are interpretive as they are based on scale scores. This study used proficiency levels (1.0-6.0 in increments of 0.1) to measure ELP.

Grouping Variables

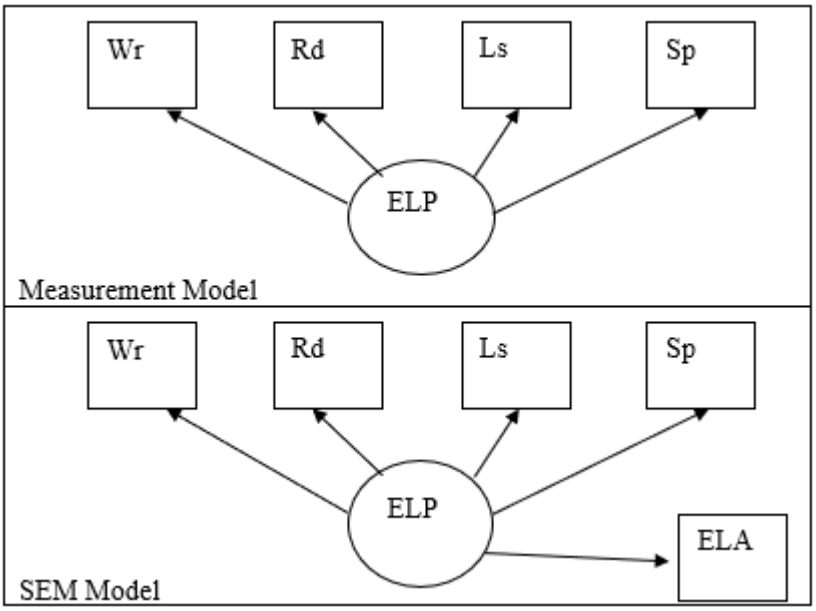
Two grouping variables were used in the data analysis, ethnicity and the years in the US schools. Since most EL students are Hispanic (McFarland et al., 2017), Hispanic and non-Hispanic EL students were compared in MI and PI. Thompson (2017) suggested that while reaching listening and speaking proficiency may happen within two years, reading and writing can take up to seven years. We used 3 years as the cut score to stay between two and seven years in this study. Thus, EL students were grouped as students who have been in the US schools for 3 years or less and students who have been in the US schools for more than three years.

Data Analyses

This study followed Culpepper et al. (2019)'s suggestion to examine MI and PI simultaneously with Multigroup Structural Equation Modeling (MGSEM). The criterion variable for all analyses was ELA assessment achievement, and the predictor variables were the four ELP scores (reading, speaking, writing, and listening). EL students' ethnicity (i.e., Hispanic and non-Hispanic) and years in the US schools (i.e., ≤ 3 years and > 3 years) were the grouping variables for examining MI and PI.

Figure 3-1

Measurement Model and SEM



Note. Rd= WIDA-reading, Wr= WIDA-writing, Ls= WIDA-listening, Sp= WIDA-speaking ELP = English language proficiency, ELA= English Language Art

Before conducting MGSEM, measurement and structural equation models (SEM) were tested for each grade level (Figure 3-1) to decide the final model to be used in MGSEM. A unidimensional measurement model for the ELP measures was tested using confirmatory factor analysis (CFA) with one-factor and four observed variables (listening, speaking, reading, and

writing). After a reasonable fit was obtained for the CFA, the SEM model was tested to examine the structural path between the ELA measures and the ELA achievement score.

Data cleaning and descriptive statistics were conducted using SPSS 26. Wicherts et al. (2005) suggested using Maximum likelihood (ML) with multigroup SEM for MI & PI, so R *lavaan* package (Rosseel, 2015) with maximum likelihood estimation was used for latent variable analysis. The following fit indices were used to examine the Goodness-of-fit: χ^2, Comparative Fit Index (CFI), and Standardized Root Mean Squared Residual (SRMR). CFI values of more than .90 (Bentler, 1990) and SRMR values of less than .10 indicated a good fit (Hu & Bentler, 1999; Kline, 2016). Although the Root Mean Square of Approximation (RMSEA) is a commonly used model fit indices, it was not used in the current study because it is sensitive to a small degree of freedom (Kenny et al., 2015), and previous literature also did not use RMSEA while examining measurement invariance and prediction validity using MGSEM (Skinner et al., 2011).

Testing for MI requires a series of steps that build upon one another and determine a baseline model for each group. Once baseline models were established, testing for invariance entailed a hierarchical set of steps that typically begins with determining a well-fitting multigroup baseline model for an ordered and increasingly restricted model. The testing procedure used the following steps: (a) null model (Figure 1), (b) test the equality of factor structure including the number of factors and factor loading patterns (configural invariance), (c) test the equality of factor loadings (weak-metric invariance), (d) test the equality of indicator intercepts (strong-scalar invariance), and (e) test equality of the regression coefficient (prediction invariance). The configural model served as the baseline model to which the subsequent models were compared. Although there are four levels of measurement invariance, configural, weak,

strong, and strict, according to Muthén and Muthén (2012), strong measurement invariance is sufficient to determine measurement invariance. Thus, measurement invariance was tested at three levels, configural, weak, and strong invariance. Lastly, the regression coefficients were constrained to be equal to examine prediction invariance between the groups.

The χ^2 difference ($\Delta\chi^2$) and the difference in CFI values (ΔCFI) were used to compare the fit between models. While insignificant χ^2 difference (p<.05) result indicates measurement invariance, it is necessary to note that it is sensitive to sample size (Kline, 2016). Considering our large sample size, ΔCFI was used to examine the measurement invariance between the models. According to Cheung and Rensvold (2002), when the ΔCFI is equal to or less than 0.01, it can be interpreted as evidence of measurement invariance among unconstrained and constrained models across the groups. The lavTestScore in the lavaan package (Rosseel, 2015) was used to determine which constrained parameters were freed to vary.

Results

Descriptive Statistics

The descriptive statistics of the observed variables are presented in Table 3-2 for each grade level. Since WIDA ACCESS scale scores are vertically scaled, scores can be compared across grades (WIDA, 2016). ELP composite scores and subscores were higher in 3rd grade compared to 8th and 10th grades. For instance, the mean ELP composite score was 3.62 and 2.79 and 2.56 for 3rd, 8th and 10th grades, respectively.

Table 3-2

Descriptive Statistics of the Variables Based on Grade Levels

	3		8		10	
Variables	Mean	SD	Mean	SD	Mean	SD
Reading Proficiency Level	3.99	1.57	3.09	1.47	3.54	1.51
Writing Proficiency Level	3.53	0.75	3.13	0.88	3.44	0.80
Listening Proficiency Level	4.78	1.52	4.63	1.41	4.06	1.42
Speaking Proficiency Level	2.82	0.79	2.79	0.94	2.56	0.79
ELA Scale Score	433.14	10.58	444.04	8.72	136.89	7.36
Student is Hispanic	.80	0.40	.80	0.40	.78	0.42
Years in the US schools is 3 years or less	.85	0.35	.38	0.49	.38	0.49

Confirmatory Factor Analysis and Structural Equation Modeling Results

Before conducting MGSEM, measurement and SEM models were tested (Figure 3-1). Table 3-3 presents the goodness-of-fit indices for these models. Single-factor CFA results suggested an appropriate model fit for all grade levels, 3rd, 8th, and 10th, based on CFI and SRMR indices. Although the χ^2 test is rejected, CFI is more than .90, and SRMR is smaller than .10 An SEM model was built based on the measurement model, and the structural model included a path from ELP to ELA achievement for prediction.

Table 3-3

Goodness-of-fit indices for measurement and SEM models

	Model	χ^2(df)	SRMR	CFI
3	Measurement model	89.75 (2) ***	.013	.995
	SEM	1400.11(5) ***	.043	.948
8	Measurement model	51.73(2) ***	.013	.995
	SEM	696.99 (5) ***	.040	.951
10	Measurement model	266.08 (2) ***	.031	.971
	SEM	288.40 (5) ***	.038	.960

Note. CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual. *p< .05; **p< .01; ***p< .001.

Multigroup Invariance Testing across Ethnicity

Goodness-of-fit indices for tests of multigroup invariance for ethnicity (Hispanic and non-Hispanic) are presented in Table 3-4. Although χ^2 was significant for all levels of invariance testing, this was expected because χ^2 is sensitive to a large sample size. The other fit indices supported a good model fit for the configural invariance, SRMR < .1 and CFI >.90. Similarly, weak, strong, and prediction invariance models did not indicate a significant difference between Hispanic and non-Hispanic students based on ΔCFI (smaller than .01) in 3rd and 8th grades. Since testing strong invariance was not retained for 10th grade (ΔCFI >.01), the intercept for listening was freed to vary based on lavTestScore (χ^2 = 43.259, p<.001). The results suggested partial strong invariance across Hispanic and non-Hispanic. The intercepts for listening were higher for Hispanics (4.33) than non-Hispanic students (4.06), suggesting that among EL students with the same level of English proficiency, Hispanic students were more likely to receive a higher listening level than non-Hispanic students. In conclusion, the results indicated that measurement invariance was attained in 3rd and 8th grades, and partial strong invariance was retained in 10th grades across Hispanic and non-Hispanic. Also, the analysis supported prediction invariance among Hispanic and non-Hispanic EL students in the 3rd, 8th, and 10th grades.

Table 3-4

Goodness-of-fit indices for Tests of Multigroup Invariance: Ethnicity

	Model	Comparative model	χ^2(df)	$\Delta\chi^2$(df)	SRMR	CFI	ΔCFI
3rd grade	1. Configural		1392.79 (10) ***	-	.037	.947	-
	2. Weak (metric)	2 vs. 1	1405.24 (13) ***	12.45 (3) *	.039	.947	.000
	3. Strong (scalar)	3 vs. 1	1503.66 (17) ***	110.87 (7) ***	.042	.943	.004
	4.prediction invariance	4 vs. 1	1509.03 (18) ***	116.24 (8) ***	.043	.943	.006
8th grade	1. Configural	-	661.97 (10) ***	-	.035	.953	-
	2. Weak (metric)	2 vs. 1	720.60 (13) ***	58.63 (3) *	.043	.949	.004

	3. Strong (scalar)	3 vs. 1	787.86 (17) ***	125.89 (7) ***	.047	.944	.009
	4.prediction invariance	4 vs. 1	842.08 (18) ***	180.11 (8) ***	.054	.940	.013
10^{th} grade	1. Configural	-	294.37 (10) ***	-	.033	.960	-
	2. Weak (metric)	2 vs. 1	311.86 (13) ***	17.49 (3) *	.040	.958	.002
	3. Strong (scalar)	3 vs. 1	376.42 (17) ***	82.05 (7) ***	.044	.949	.011
	3.1 Intercept for listening freed to vary	3.1 vs. 1	332.07 (16) ***	37.70 (6) ***	.042	.955	.005
	4.Prediction invariance	4 vs. 1	339.95 (17) ***	45.58 (7) ***	.045	.954	.006

Note. CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual. *p< .05; **p< .01; ***p< .001. 2 vs. 1 indicates the model comparison between the configural invariance model and the weak invariance model.

Multigroup Invariance Testing across Years in the US Schools

3rd Grade

Multigroup invariance testing was conducted across years in the US schools for two groups, 3 years/less and more than 3 years for each grade level. Table 3-5 presents the goodness-of-fit indices for tests of multigroup invariance in 3rd grade across years in US schools. Similar to multigroup invariance testing across ethnicity, configural invariance was supported by the SRMR and CFI fit indices, which means that a factor structure is the same across the years in the US schools for all three grade levels. The configural, weak, strong, and prediction invariances were retained for 3rd grade based on ΔCFI (smaller than .01). These results indicated measurement and prediction invariance across years in the US schools for 3rd grades. Further analysis examined the MI and PI for the 8th and 10th grades.

Table 3-5

Goodness-of-fit indices for Tests of Multigroup Invariance: Years in the US Schools – 3rd Grade

Model	Comparative model	χ^2(df)	$\Delta\chi^2$(df)	SRMR	CFI	ΔCFI
1. Configural	-	1335.69 (10) ***	-	.036	.950	-
2. Weak (metric)	2 vs. 1	1375.14 (13) ***	39.45 (3) ***	.040	.949	.001
3. Strong (scalar)	3 vs. 1	1567.52 (17) ***	231.83 (7) ***	.045	.942	.008
4.Prediction invariance	4 vs. 1	1571.83 (21) ***	236.14 (8) ***	.046	.941	.009

Note. CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual. *p< .05; **p< .01; ***p< .001. 2 vs. 1 indicates the model comparison between the configural invariance model and the weak invariance model.

8th Grade

Table 3-6 presents the goodness-of-fit indices for tests of multigroup invariance in 8th grade across years in US schools. While configural measurement invariance was retained for 8th grade based on ΔCFI (smaller than .01), imposing equality constraints on factor loadings rejected the weak measurement invariance model according to ΔCFI (larger than .01). The factor loading for reading was freed to vary in the weak invariance model according to the lavTestScore function (χ^2 = 156.914, p<.001). Although this change improved the model fit, further modification was suggested to free a factor loading for speaking (χ^2 = 16.030, p<.001). After relaxing the factor loading constraints for reading and speaking, the model presented a non-significant difference with the baseline model as ΔCFI being smaller than .01. Testing strong invariance by freeing reading and speaking factor loadings resulted in measurement noninvariance based on ΔCFI (larger than .01) so reading (χ^2 =144.454, p<.001), writing (χ^2 =15.469, p<.001), and listening (χ^2 =56.189, p<.001) intercepts were freed to vary in this order as suggested by lavTestScore function. Since freeing reading, listening, and writing intercepts did not improve the model fit (ΔCFI > .01), it was concluded that strong invariance was not met

for 8th graders. Testing prediction invariance by freeing reading and speaking factor loadings and all intercepts indicated prediction invariance (ΔCFI < .01).

Table 3-6

Goodness-of-fit indices for Tests of Multigroup Invariance: Years in the US Schools- 8th Grade

Model	Comparative model	χ^2(df)	$\Delta\chi^2$(df)	SRMR	CFI	ΔCFI
1. Configural	-	495.28 (10) ***	-	.028	.963	-
2. Weak (metric)	2 vs. 1	688.84 (13) ***	193.65 (3) ***	.058	.948	.015
2.1 Factor loading for reading freed to vary	2.1 vs 1	511.20 (12) ***	15.93 (2) ***	.033	.962	.001
2.2 Factor loadings for reading and speaking freed to vary	2.2 vs 1	495.58 (11) ***	.31 (1)	.028	.963	.000
3. Strong (scalar) Factor loadings for reading and speaking freed to vary	3 vs. 1	924.47 (15) ***	429.20 (5) ***	.067	.930	.033
3.1 Strong (scalar) Factor loadings for reading, speaking and intercept for reading freed to vary	3.1 vs. 1	773.85 (14) ***	278.58 (4) ***	.055	.942	.021
3.2 Strong (scalar) Factor loadings for reading, speaking and intercepts for reading and writing freed to vary	3.2 vs. 1	758.01 (13) ***	262.74 (3) ***	.053	.943	.020
3.3 Strong (scalar) Factor loadings for reading, speaking and intercepts for reading, writing and listening freed to vary	3.3 vs. 1	698.77 (12) ***	203.49 (2) ***	.048	.947	.016
4.Prediction invariance Factor loadings for reading, speaking and all intercepts are freed to vary	4 vs. 1	542.02 (12) ***	46.75 (2) ***	.038	.959	.004

Note. CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual. *p< .05; **p< .01; ***p< .001. 2 vs. 1 indicates the model comparison between the configural invariance model and the weak invariance model.

Table 3-7 presents the standardized and unstandardized regression coefficients for two groups for the prediction invariance model where factor loadings for reading, speaking, and all intercepts were freed to vary. The students who have been in the US schools for more than three

years have higher reading unstandardized coefficient (1.96) than the students who have been in the US school for less than 3 years (1.49). When examining prediction invariance by constraining the regression coefficients, there was a non-significant difference with the baseline model. Thus, the results suggested the measurement invariance was not attained, but there was a prediction invariance across years in the US schools for 8th graders.

Table 3-7

Unstandardized and Standardized Coefficients and Intercepts of MGSEM – 8th Grade

	3 years or less in the US schools			More than 3 years in the US schools		
	Estimate	SE.	Std. Estimate	Estimate	SE.	Std. Estimate
Factor loadings						
ELP → Wr	1.00		.92	1.00		.81
ELP→ Rd	1.49	.03	.87	1.96	.04	.81
ELP→ Ls	1.42	.02	.84	1.42	.02	.69
ELP→ Sp	.91	.02	.82	.75	.03	.52
ELA → ELP	8.69	.15	.80	8.69	.15	.63
Intercepts						
Wr	2.75	.02	2.77	3.37	.01	4.77
Rd	2.82	.04	1.79	3.26	.03	2.37
Ls	4.07	.04	2.63	4.98	.02	4.22
Sp	2.44	.02	2.41	3.01	.02	3.67
ELA	443.22	.23	44.84	444.59	.14	56.67

10th Grade

Table 3-8 presents the goodness-of-fit indices for tests of multigroup invariance in 10th grade across years in US schools. Although configural measurement invariance was retained for 10th grade based on ΔCFI (smaller than .01), imposing equality constraints on factor loadings rejected the weak measurement invariance model according to ΔCFI (larger than .01). The factor loading for reading (χ^2 =66.001, p<.001) was freed to vary in the weak invariance model according to the lavTestScore function. Although this change improved the model fit, further modification was suggested to free the factor loading for listening (χ^2 =21.590, p < .001). Once

freeing the factor loading constraints for reading and listening, the model presented a non-significant difference with the baseline model (ΔCFI < .01). Then strong invariance was tested by relaxing reading and listening factor loadings resulted in measurement variance (ΔCFI >.01). Therefore, listening (χ^2 =58.109, *p*< .001) and writing (χ^2 =69.763, *p* <.001) intercepts were freed to vary in that order according lavTestScore function. Since freeing listening, and writing intercepts improved the model fit (ΔCFI < .01), it was concluded that partial strong invariance was met for 10th graders.

Table 3-8

Goodness-of-fit indices for Tests of Multigroup Invariance: Years in the US Schools – 10th Grade

Model	Comparative model	χ^2(df)	$\Delta\chi^2$(df)	SRMR	CFI	ΔCFI
1. Configural	-	296.30 (10) ***	-	.035	.955	-
2. Weak (metric)	2 vs. 1	396.19 (13) ***	99.90 (3) ***	.060	.940	.015
2.1 Factor loading for reading freed to vary	2.1 vs. 1	324.01 (12) ***	27.71 (2) ***	.044	.951	.004
2.2 Factor loadings for reading and listening freed to vary	2.2 vs. 1	301.11 (11) ***	4.81 (1) *	.038	.955	.000
3. Strong (scalar)	3 vs. 1	447.40 (15) ***	151.1 (5) ***	.057	.933	.022
3.1. Factor loadings for reading, listening and the intercept for listening freed to vary	3.1 vs. 1	389.95. (14) ***	93.65 (4) ***	.051	.941	.014
3.2. Factor loadings for reading, listening and the intercepts for listening and writing freed to vary	3.2 vs. 1	320.20 (13) ***	23.90 (3) ***	.040	.952	.003
4.prediction invariance Factor loadings for reading, listening and the intercepts for listening and writing freed to vary	4 vs. 1	379.27 (14) ***	82.98 (4) ***	.053	.943	.012

Note. CFI = comparative fit index; SRMR = Standardized Root Mean Squared Residual. **p*< .05; ***p*< .01; ****p*< .001. 2 vs. 1 indicates the model comparison between the configural invariance model and the weak invariance model.

Table 3-9 presents the standardized and unstandardized regression coefficients across two groups. Like 8th graders, students who have been in the US schools for more than three years have higher reading unstandardized coefficient (3.04) than students who have been in the US school less than 3 years (1.95) for 10th graders. When examining prediction invariance by constraining the regression coefficient, there was a significant difference with the baseline model based on ΔCFI (larger than .01). Thus, the results suggested a partial strong measurement invariance was attained but there is no prediction invariance across years in the US schools for 10th graders. The students who have been in the US schools for more than three years have a higher ELA unstandardized coefficient (12.39) than the students who have been in the US school less than 3 years (8.23) for 10th graders.

To sum up, the analysis indicated measurement invariance and prediction invariance according to the years in the US schools by 3rd graders. In contrast, only partial weak invariance and partial strong invariance were met by 8th and 10th graders, respectively. While there was a prediction invariance across the years in the US schools for 8th, there was no prediction invariance among 10th graders.

Table 3-9

Unstandardized and Standardized Coefficients and Intercepts of MGSEM – 10th Grade

	3 years or less in the US schools			More than 3 years in the US schools		
	Estimate	SE.	Std. Estimate	Estimate	SE.	Std. Estimate
Factor loadings						
ELP → Wr	1.00		.76	1.00		.60
ELP → Rd	1.95	.06	.88	3.04	.11	.81
ELP → Ls	1.79	.06	.85	2.27	.09	.69
ELP → Sp	.95	.03	.77	.95	.03	.51
ELA → ELP	8.23	.29	.75	12.39	.51	.68
Intercepts						
Wr	3.46	.03	3.98	3.68	.02	5.66
Rd	3.84	.03	2.60	3.84	.03	2.65
Ls	4.10	.04	2.95	4.49	.03	3.55

Sp	2.70	.02	3.30	2.70	.02	3.72
ELA	138.19	.15	18.98	138.19	.15	19.41

The Difference among Latent Score and Observe Score Prediction of ELA Achievement

We used latent scores to examine MI and PI in MGSEM analysis. In this section, a series of regression analyses were conducted to understand the relationship between ELA achievement, ELP and group memberships, ethnicity, and the years in the US schools based on observed scores (Table 3-10). The regression results indicated that Hispanic students received an ELA score of .06 and .04 standard deviation units less than non-Hispanic with the same language proficiency in 3rd and 8th grades. However, there was no statistically significant difference among Hispanic and non-Hispanic students regarding ELA achievement in 10th grades. While regression analysis indicated a significant difference between Hispanic and non-Hispanic EL students in terms of ELA achievement in 3rd and 8th grades, latent score analysis concluded that there was measurement and prediction invariance across ethnic groups. There was a partial strong measurement invariance between Hispanic and non-Hispanic in 10th grade, although the regression results suggested no significant difference between Hispanic and non-Hispanic.

The regression analyses indicated that the students, who have been in the US schools three years or less, received ELA scores .06, .14, and .08 standard deviation units higher than students who have been in the US for more that 3 years in 3rd, 8th and 10th respectively. The highest difference was in 8th grades between students who have been in the US schools three years or less and students who have been in the US schools more than three years. Although regression results indicated a significant difference across years in the US schools among 8th graders, latent score analysis suggested a partial weak measurement invariance rather than prediction noninvariance. This result could mean that the observed score difference in ELA scores across the years in the US schools could be measurement noninvariance for 8th graders.

The regression results indicated that the students, who have been in the US schools three years or less, received a .08 standard deviation unit higher ELA score than the students who have been in the US schools for more than three years with the same ELP score in 10th grade. Similar to the regression analysis results, which suggested the significant difference across years in the US schools in 10th grade, latent score analysis indicated a partial strong measurement invariance and prediction noninvariance.

Lastly, the regression results showed that R^2 values decrease as grades increase. For instance, ELP and Hispanic variables explained 64%, 48%, and 43% variance in ELA achievement in 3rd, 8th and 10th grades, respectively. Likewise, predictors of ELP and the years in the US schools explained 64%, 49%, and 43% variance in ELA achievement in 3rd, 8th, and 10th grades. Besides, ELP's standardized regression coefficients are decreasing in the same pattern. Therefore, these results may suggest that the amount variance explained by ELP in ELA achievement went down from 3rd grade to 10th grade.

Table 3-10

Regression Analysis Summary for Predicting ELA Achievement

Grade	Variable	B	SE.	β	F	Adjusted R^2
3	Constant	401.71	.31		F(2)=8633.84***	.64
	ELP observed score	9.04	.07	.79***		
	Hispanic	-1.56	.16	-.06***		
8	Constant	425.95	.37		F(2)=2258.05***	.48
	ELP observed score	5.92	.09	.68***		
	Hispanic	-.96	.23	-.04***		
10	Constant	119.40	.43		F(2)=1167.62***	.43
	ELP observed score	5.34	.11	.66***		
	Hispanic	-.13	.24	-.007		
3	Constant	398.79	.29		F(2)=8649.68***	.64
	ELP observed score	9.06	.07	.79***		
	3 years or less in the US schools	1.85	.18	.06***		
8	Constant	422.99	.33		F(2)=2396.61***	.49
	ELP observed score	6.32	.09	.73***		
	3 years or less in the US schools	2.43	.19	.14***		

10	Constant	118.07	.44		F(2)=1192.29***	.43
	ELP observed score	5.56	.12	.68***		
	3 years or less in the US schools	1.17	.22	.08***		

Note. *p< .05; **p< .01; ***p< .001.

Discussion

This study examined the MI and PI among EL students focusing on ethnicity and the years in the US schools. The latent score analysis suggested measurement and prediction invariances was retained between Hispanic and non-Hispanic EL students in 3rd and 8th grades, but there was a partial strong measurement variance in 10th grade. Previous research indicated that reaching English proficiency can take longer for Hispanic EL students (Slama, 2014; Thompson, 2017), and in this study, we examined if measurement and prediction invariance could explain the source of this difference between Hispanic and non-Hispanic students. The results of MGSEM analysis suggested that MI and PI do not seem to be the reason for this difference, although the strong measurement invariance was only partially met for 10th graders between Hispanic and non-Hispanic students. On the other hand, these results could be interpreted as additional validity evidence for the WIDA ACCESS assessment since there is a uniform interpretation of scores between Hispanic and non-Hispanic EL students.

LTEL students have been a concerning issue regarding EL education (Callahan, 2005; Slama, 2014; Flores et al., 2012). This study examined MI and PI among the years in the US schools. The results presented a prediction invariance across years in the US schools for 3rd, and 8th graders. However, prediction invariance was not retained among 10th graders, and there was a partial weak measurement invariance attained for 8th and 10th graders. The results presented that the students who have been in the US schools for more than 3 years tended to receive higher scores on ELP subscores. These results are important for LTEL because they presented a

potential source of error. Besides, conducting the same MGSEM for multiple grades concluded different results which could raise the concern of longitudinal MI (Kline, 2016).

One may argue that students' ELP increases as their grade level is goes up, and ELP could be the reason for MI across the years in the US schools for 8th and 10th grades. However, the regression results suggested that the effect of ELP on ELA achievement decreases as grade level increases. On the other hand, it is important to note that the WIDA ACCESS assessment aims to measure English academic proficiency. Therefore, the findings of this study should be a steppingstone to analyze further validity evidence to address the heterogeneity among EL students in terms of ethnicity, time in the US, and first language proficiency.

The regression analyses indicated that the percentage of variance explained by English proficiency in ELA achievement decreases as grade level goes up. A possible explanation could be there are other factors to influence ELA achievement. For example, EL students may experience anxiety due to learning in an English-speaking classroom and dealing with stigmas (Orosco & Klingner, 2010). Thus, according to second language instruction principles, facilitating EL students to interact with their peers through speaking, reading, writing, and listening in their second language could improve their English proficiency (Ellis, 2005).

Fairness in Testing

The lack of MI and PI have implications for measurement and interpretations of ELP assessment as it means the scores cannot be compared across the years in the US schools. Our analysis showed a lack of strong measurement invariance among the years in the US school for 8th graders. This measurement non-invariance means that observed scores may not be accurately compared across students who have been in the US schools less than three years and students who have been in the US schools for more than three years.

This study addressed the fairness issues among the EL population by examining measurement and prediction invariance using MGSEM. However, the results should be interpreted with caution. The dimensionality of language efficacy was not examined in this study because the item-level data was not available. In addition, dimensionality was a potential cause of measurement invariance. Second, the latent score and regression results were based on real data, and future studies could examine this comparison with simulation studies that would allow researchers to know the true MI and PI being tested. Another limitation is that EL students who did not take the ELA assessment were excluded from this study.

Future research should examine MI and PI using item-level data among the EL population and investigate MI for ELA achievement between EL and non-EL students because MI is necessary for accurate observe score comparison. In this study, we examined the MI and PI for multiple grades based on ethnicity and the year in the US schools. Future studies may investigate MI over time with the same cohort among EL students.

References

Abedi, J., & Linquanti, R. (2012, January). *Issues and opportunities in improving the quality of large scale assessment systems for ELLs*. Paper presented the Understanding Language conference, Palo Alto, CA. Retrieved from http://ell.stanford.edu/publication/issues-and-opportunities-improving-quality-large-scale-assessment-systems-ells

Abedi, J. (2007). *English language proficiency assessment in the nation: Current status and future practice*. Davis, CA: University of California.

Abedi, J. (2006a). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record*, 108(11), 2282.

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2016). Differential prediction generalization in college admissions testing. *Journal of Educational Psychology, 108,* 1045–1059.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Berry, C. M., & Zhao, P. (2015). Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance. *Journal of Applied Psychology, 100,* 162–179.

Brooks, M. D. (2018). Pushing past myths: Designing instruction for long-term English learners. *TESOL Quarterly, 52* (1), 221–233.

Bryant, D. (2004). *The effects of differential item functioning on predictive bias*. Unpublished doctoral dissertation), University of Central Florida, Orlando, Florida.

Bunch, M. B. (2011). Testing English language learners under No Child Left Behind. *Language Testing,* 28, 323–341. doi:10.1177/026553221140418.

Butler, F. A., Stevens, R., & Castellon, M. (2007). ELLs and standardized assessments: The interaction between language proficiency and performance on standardized tests. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 27–49). New Haven, CT: Yale University Press.

Calderón, M., Slavin, R. E., & Sánchez, M. (2011). Effective instruction for English learners. *Future of Children, 21,* 103–127.

Callahan, R. M. (2005). Tracking and high school English learners: Limited opportunities to learn. *American Educational Research Journal,* 42, 305–328. doi:10.3102/00028312042002305

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233-255.

Clausen, D. (2017). Empowering English Language Learners in Postsecondary Classrooms An Inquiry into Best Practices. *Journal of Alternative Perspectives in the Social Sciences*, *8*(4).

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.

Corcoran, R. P. (2018). An embodied cognition approach to enhancing reading achievement in New York City public schools: Promising evidence. *Teaching and teacher education, 71*, 78-85. https://doi.org/10.1016/j.tate.2017.11.010

Culpepper, S. A., Aguinis, H., Kern, J. L., & Millsap, R. (2019). High-stakes testing case study: A latent variable approach for assessing measurement and prediction invariance. *Psychometrika*, 84(1), 285-309.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, England: Multilingual Matters.

Ellis, R. (2005). *Instructed second language acquisition: A literature review*. Wellington, Australia: Research Division, Ministry of Education.

Every Student Succeeds Act. (2015). *Every Student Succeeds Act of 2015*, Pub. L. No 114–95 (2015-2016).

Fischer, F. T., Schult, J., & Hell, B. (2013). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational Psychology, 105*, 478–488.

Flores, S. M., Batalova, J., & Fix, M. (2012). *The educational trajectories of English language learners in Texas.* Washington, DC: Migration Policy Institute.

Flores, N., Kleyn, T., & Menken, K. (2015). Looking holistically in a climate of partiality: Identities of students labeled long-term English language learners. *Journal of Language, Identity and Education*, 14, 113–132. doi:10.1080/15348458.2015.1019787

Flores, S. M., Park, T. J., Viano, S. L., & Coca, V. M. (2017). State policy and the educational outcomes of English learner and immigrant students: Three administrative data stories. *American Behavioral Scientist, 61*, 1824–1844. doi:10.1177/0002764217744836

Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center.

Fry, R., & Gonzales, F. (2008). One-in-five and growing fast: A profile of Hispanic public-school **No table of figures entries found.**students. Washington, DC: Pew Hispanic Center. Retrieved from http://files.eric.ed.gov/fulltext/ED502556.pdf

Grant, R. & Kraninger, N. (2019). *ACCESS for ELLs 2.0 Construct Validation Study - ACCESS for ELLs 2.0, Series 400 with 2015-2016 Data*. WIDA Technical Report TR-2019-1 (unpublished).

He, P. (2021). *ELL Proficiency Status and ELA Scores of ELL Students across Grade Levels and Gender: A Quantitative Nonexperimental Study* (Doctoral dissertation, Northcentral University).

Heise, M. (2017). From no child left behind to every student succeeds: Back to a future for education federalism. *Colum. L. Rev.*, *117*, 1859. https://doi.org/10.31228/osf.io/kdfje

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. HTTPS:// doi.org/10.1080/10705519909540118.

Humphreys, L. G. (1952). Individual differences. *Annual Review of Psychology, 3*, 131–150.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*, 131–161.

Johnson, J. L., Trantham, P., & Usher-Tate, B. J. (2019). An Evaluative Framework for Reviewing Fairness Standards and Practices in Educational Tests. *Educational Measurement: Issues and Practice, 38*(3), 6-19.

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology, 101*, 569–581.

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., Wilkin- son-Flicker, S., & Kristapovich, P. (2016). *The Condition of Education 2014* (NCES 2014–083). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http:// nces.ed.gov/pubsearch

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological methods & research*, *44*(3), 486-507. DOI: 10.1177/0049124114543236

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford publications. (Fourth Edition).

Krogstad, J. M. (2016, April 20). Rise in English proficiency among U.S. Hispanics is drive by the young. https://www.pewresearch.org/fact-tank/2016/04/20/rise-in-english-proficiency-among-u-s-hispanics-is-driven-by-the-young/

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31*(1), 3-12.

Lopez, M. H., Krogstad, J. M., & Flores, A. (2018, April 2). Most Hispanic parents speak Spanish to their children, but this is less the case in later immigrant generations. https://www.pewresearch.org/fact-tank/2016/04/20/rise-in-english-proficiency-among-u-s-hispanics-is-driven-by-the-young/

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Charlotte: Information Age Publishing Inc.

Mattern, K. D., & Patterson, B. F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, and Pierce (2010). *Journal of Applied Psychology, 98*, 134–147.

McARdle, F., Grieshaber, S., & Sumsion, J. (2019). Play meets early childhood teachers education. *The Australian Educational Researcher: A Publication of the Australian Association for Research in Education,* 46(1),155. https://doi.org/10.1007/s13384-018-0293-8

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., and Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017- 144). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144.

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7,* 361–388. doi:10.1177/1094428104268027

Miley, S. K., & Farmer, A. (2017). English Language Proficiency and Content Assessment Performance: A Comparison of English Learners and Native English Speakers Achievement. *English Language Teaching*, *10*(9), 198-207.

Millsap, R. E. (2007). Invariance in measurement and prediction revisited. *Psychometrika*, *72*(4), 461-473.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user guide (version 7)*. Los Angeles: Muthén & Muthén.

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2018). *English learners in STEM subjects: Transforming classrooms, schools, and lives*. Washington, DC: National Academies Press.

National Center for Education Statistics. (2014). *Public high school graduation rates, 2011–12.* Retrieved from http://nces.ed.gov/programs/coe/indicator_coi.asp

National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs. (2007). Retrieved from http://sbo.nn.k12.va.us/esl/documents/ncela_fast_faqs.pdf

No Child Left Behind Act of 2001, PL 107-110, 20 USC. (2002).

Orosco, M. J., & Klingner, J. (2010). One school's implementation of RTI with English language learners: "Referring into RTI". *Journal of Learning Disabilities*, 43, 269–288.

Parker, C. E., Louie, J., & O'Dwyer, L. (2009). New Measures of English Language Proficiency and Their Relationship to Performance on Large-Scale Content Assessments. Issues & Answers. REL 2009-No. 066. *Regional Educational Laboratory Northeast & Islands*.

Polat, N., Zarecky-Hodge, A., & Schreiber, J. B. (2016). Academic growth trajectories of ELLs in NAEP data: The case of fourth- and eighth-grade ELLs and non-ELLs on Mathematics and Reading tests. *Journal of Educational Research, 109*(5), 541-553. doi:10.1080/00220671.2014.993461

Portes, A., & Hao, L. (1998). E pluribus unum: Bilingualism and loss of language in second generation. *Sociology of Education*, *71*(4), 269–294.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41,* 71-90. http://dx.doi.org/10.1016/j.dr.2016.06.004

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., ... & Barendse, M. (2015). Package "lavaan": Latent variable analysis. R Package. Available online at: http://lavaan.org.

Skinner, M. L., MacKenzie, E. P., Haggerty, K. P., Hill, K. G., & Roberson, K. C. (2011). Observed parenting behavior with teens: Measurement invariance and predictive validity across race. *Cultural Diversity and Ethnic Minority Psychology, 17*(3), 252. doi:10.1037/a0024730.

Slama, R. B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology, 104*(2), 265.

Slama, R. B. (2014). Investigating whether and when English learners are reclassified into mainstream classrooms in the United States: A discrete-time survival analysis. *American Educational Research Journal, 51,* 220-252.

Thompson, K. D. (2017). English learners' time to reclassification: An analysis. *Educational Policy, 31*(3), 330-363.

Thurlow, M. L., Wu, Y.-C., Lazarus, S. S., & Ysseldyke, J. E. (2016). Special education--non-special education achievement gap in math: Effects of reporting methods, analytical techniques, and reclassification. *Exceptionality, 24*(1), 32–44.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89,* 696–716.

WIDA Consortium. (2012). *2012 amplification of the English Language Development Standards Kindergarten–Grade 12*. Madison, WI: Board of Regents of the University of Wisconsin System.

WIDA Consortium. (2020). *Annual Technical Report for ACCESS for ELLs Online English Language Proficiency Test Series 403, 2018–2019 Administration*. Madison, WI: Board of Regents of the University of Wisconsin System.

WIDA (2016). *ACCESS for ELLs 2.0 Interpretive Guide for Score Reports Kindergarten–Grade 12.* Madison, WI: Board of Regents of the University of Wisconsin System.

WIDA (2020) *Access Tests*. The Board of Regents of the University of Wisconsin System. Retrieved from https://wida.wisc.edu/assess/access/tests

Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P., Chang, S. M., & Farnsworth, T. (2008b). *Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review* (CRESST Technical Report No. 732). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Wright, W. (2006, November). Acatch-22 for language learners. Educational Leadership, 64(3), 22-27. Retrieved from http://search.ebscohost.com

OVERALL CONCLUSION

This dissertation aims to examine fairness issues regarding English learners (EL) in US schools. This dissertation aims (a) to summarize EL test accommodations research on computer-based testing (CBT) by conducting a meta-analysis; (b) to predict EL with disabilities postschool outcomes from transition planning, parent expectation, and adaptive behaviors; and (c) to examine MI and PI simultaneously based on ethnicity, and the years in the US schools using EL large-scale assessment data.

The first article analyzed the validity and the effectiveness of EL test accommodations on CBT. Five studies examined the validity of EL test accommodations on CBT, and these studies included 10 effect sizes and 2779 non-EL students. The total effect size for the validity of EL test accommodations on CBT was -0.02 *SD* (p=.88). This result revealed that EL test accommodations on CBT did not influence non-EL students' test scores, which suggests that the EL test accommodations did not change the constructs being measured in the test. In total, eight studies examined the effectiveness of EL test accommodations on CBT, and these studies included 28 effect sizes and 5987 EL students. The total effect size of EL test accommodations on CBT was .12 *SD* (p <.05). The results indicated that EL students who used test accommodations on CBT improved by .12 standard deviation units.

The second article investigated the relationships between EL with disabilities' postschool outcomes (PSO) with adaptive behaviors, parent expectations, and transition planning using the National Longitudinal Transition Study-2 data. The transition directly affected PSO, with a one standard deviation increase in transition planning resulting in a .96 increase in PSO. The model explained 85% of the variance in parent expectations, 88% of the variance in transition planning, and 92% of PSO variance. The results suggested that transition planning mediated the effects

between PSO and adaptive behaviors and parent expectations. This study expanded the literature regarding EL with disabilities' PSO and the importance of parent expectations and transition planning.

The third article studied the measurement and prediction invariance simultaneously based on latent scores and compare the results with observed scores prediction using multigroup structural equation modeling across ethnicity and the years in the US schools for 3rd, 8th, and 10th graders. Data for this study came from a southeastern state ELA accountability assessment and WIDA's ACCESS ELP assessment in 2019. There were measurement and prediction invariances among Hispanic and non-Hispanic EL students except a partial strong invariance in 10th grade. While measurement invariance was attained for 3rd graders, only partial weak and a partial strong invariance were retained for 8th and 10th grades, respectively. There was a prediction invariance across the years in the US schools for 3rd and 8th graders but not for 10th graders. While regression analysis indicated a significant difference between Hispanic and non-Hispanic EL students in terms of ELA achievement in 3rd and 8th grades, latent score analysis concluded no measurement and prediction invariance across ethnic groups.

This dissertation contributed to the extant literature by conducting EL test accommodation meta-analysis to CBT context. The results of this study confirmed the previous meta-analysis on EL test accommodations (Li & Suen, 2012; Rios et al., 2020) and added literature that CBT accommodations were more effective for K-6 compared to grades 7-12. This study expanded the literature by analyzing EL with disabilities' postschool transition as Trainor et al. (2020) indicated that less is known about supporting students with disabilities with complex needs in the postschool transition process. Lastly, this research contributed to the

existing literature by examining MI and PI for EL large-scale assessment to demonstrate uniform interpretations of test results across ethnicity and the years in the US schools.

There are some implications for educators and policymakers in EL education. The alignment of EL instruction and accommodations is necessary for the effective use of EL test accommodations. While there was a small effect found for use of accommodations, educators need to find methods of increasing the effectiveness of accommodations. Results of the second study indicate that transition planning serves to mediate the relationship between PSO and adaptive behavior and parent expectations. Policymakers should examine which transition components strengthen the indirect relations to PSO. Thus, educators should increase the collaboration between student school and family in order to support EL with disabilities for a successful postschool transition. Third study presented some evidence of measurement non-invariance which means that observed scores may not be accurately compared across students who have been in the US schools less than three years and students who have been in the US schools for more than three years. Thus, educators and policymakers should consider the potential implications of MI in EL large-scale assessment.

This dissertation had implications for researchers. First, it has been found that EL accommodation on CBT has a small and significant effect on improving EL achievement. However, more experimental research is needed to examine the effect of frequent test accommodations on EL academic achievement. Second, future research should examine the personal characteristics (i.e., self-determination, autonomy, and self-realization) on EL postschool transition. EL's large-scale assessment indicated the lack of MI across the years in the US school for 8^{th} and 10^{th} grades and no PI among 10^{th} graders. Future research should explore MI and PI with item-level data and confirm results with simulation studies.

References

Li, H., & Suen, H. K. (2012a). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education*, 25(4), 327-346.

Rios, J. A., Ihlenfeldt, S. D., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence‐based? A multistudy systematic review and meta‐analysis. *Educational Measurement: Issues and Practice*.

Trainor, A. A., Carter, E. W., Karpur, A., Martin, J. E., Mazzotti, V. L., Morningstar, M. E., Newman, L., & Rojewski, J. W. (2020). A framework for research in transition: Identifying important areas and intersections for future study. *Career Development and Transition for Exceptional Individuals*, *43*(1), 5-17. https://doi.org/10.1177%2F2165143419864551