Examining the Impact of the COVID-19 Pandemic on English Learners’ Proficiency and Disparities Within EL Subgroups

by

Narék Sahakyan

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Educational Leadership & Policy Analysis)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 4/25/2024

The dissertation is approved by the following members of the Final Oral Committee:
Anjalé Welton, co-Chair, Professor, Educational Leadership & Policy Analysis
Christopher Saldaña, co-Chair, Assistant Professor, Educational Leadership & Policy Analysis
Xueli Wang, Professor, Educational Leadership & Policy Analysis
Courtney Bell, Professor, Learning Sciences, WCER Director
Timothy Boals, Senior Research Scientist, WIDA Executive Director and Founder
H. Gary Cook, Senior Research Fellow, Wisconsin Center for Education Research

DEDICATION

I dedicate this work to teachers.

Growing up in a family of hard-working, dedicated, and selfless professionals who devoted their life to teaching, I have had a good vantage point on how difficult and taxing it can be. Trying to educate children under severe constraints and circumstances that are usually well outside of their control, teachers are tasked with the thankless and responsible task of shaping the ideas and perspectives of future generations. The tremendous impact teachers can have on students', and especially on young emergent multilingual learners' education cannot be overstated; yet it is missing from this work, too: it is so difficult to quantify the invaluable.

From '*comrade'* Gevorgian in my elementary school, to my mother – a teacher of English, my aunt – a teacher of French, my grandparents – teachers of Mathematics and Russian, to the many teachers of multilingual learners and other students, thank you for the important and impactful work you do!

ACKNOWLEDGMENTS

This work cannot be presented without first acknowledging the deep extent of privilege that an academic researcher enjoys sitting in an ivory tower (11th floor, Wisconsin Center for Education Research) at a desk, theorizing, reading about, and analyzing data on real students' scores, disparities in achievement, and the impact of a global pandemic on these scores and disparities. From the students that take the annual assessment of their English proficiency, to their teachers, parents, peers, and school staff that help develop and shape it, and to test administrators, item and content developers, database administrators, psychometricians, applied linguists, data managers and custodians, analysts, project managers, graduate students, professors and researchers – millions of people have contributed to – WIDA's ACCESS for ELLs assessment that provides the universe of data underlying this study. This work would not have been possible without their joint efforts.

This dissertation would not be possible without the tremendous and unique contributions of each of the members of this highly esteemed dissertation committee. With their encouraging, yet always challenging ways, the faculty from ELPA – co-chairs Dr. Anjalé Welton and Dr. Christopher Saldaña, along with Dr. Xueli Wang, have given me a first-hand education in what Educational Leadership, as well as Policy Analysis should be about. I would never be in this position without your patience, understanding, and encouragement. I am also indebted to Dr. Courtney Bell, the director of the Wisconsin Center for Education Research. Thank you for stepping in to help when I needed it, despite the pressing timelines and your numerous executive duties. Your amiable

approach, quick responses, and on-point feedback, reassured me and supported this work in important ways. It is a privilege to have you on this committee.

Dr. Timothy Boals is the founder and executive director of WIDA. I could stop there, but unbelievably, you have done so much more! It has been an absolute privilege and learning experience watching you build and steer a ship as large and complex as the WIDA Consortium in turbulent and shark-infested waters, full of political icebergs and bureaucratic red-tape. Tim, thank you for your patience as I took my time to get here.

Dr. H. Gary Cook. Your key role in making sure that the data underlying this analysis would be collected, stored, maintained, and examined for research and assessment innovation purposes enabled this analysis. From the day you hired me 14 years ago you have been the best supervisor, tutor, and mentor I could ask for. The wisdom you shared with me in the thousands of meetings discussing various statistical models or approaches dwarfs even this dissertation, which in turn would not be possible without you teaching me about the countless nuances of English learner assessment and education policy world, and the intricacies of ACCESS and psychometric evaluation. Thank you for showing what true leadership and integrity, effective management, and multidisciplinary professionalism should be about. Admittedly your lessons in snorkeling and body-surfing were also instrumental in forging our research partnership and taking WIDA's technical assistance to new levels. It may not be Friday, but still, '*pau hana*'!

Dani, thank you for being an amazing friend and a brilliant thought (as well as tennis) partner throughout the years. The incredible quality and quantity of your ideas, questions, feedback, and suggestions, coupled with your infinite curiosity and willingness to help (attempting to decompose convoluted, multi-subject, and multi-line sentences like

this one, among many other things) uplifted this work to new heights; unfortunately, you don’t get to do that in this section, and fortunately the readers now know what you saved them from. Your perspectives, comments, and questions give me hope that quantitative and qualitative research can align closer together to produce more rigorous and informative mixed methods studies.

I am also very grateful to Glenn Poole, my graduate assistant, ELPA classmate and fellow dissertator, co-author, friend, and hopefully soon a colleague. The hours you invested doing detailed and rigorous, yet mundane and taxing work on cleaning data and organizing datasets, writing pages of programming code, creating, updating, and reverting variable definitions, and debating policy, theory, and statistics, made this work possible within otherwise infeasible timelines. It has been a joyous, and at times jealous experience watching you master STATA, statistical models, and quantitative concepts in record (and half-) time. Thank you for being the best assistant and a mindful partner in analysis and research. Wherever the future takes you, I anticipate plenty of collaboration and research. And now, with this project winding down, I hope to play more chess.

After almost failing statistics during my time as an undergraduate in Armenia, I rediscovered and developed a passion for examining probabilistic concepts and exploring statistical relationships between various factors during my time at the Agricultural and Applied Economics Department. My very first graduate adviser, great mentor, and soon a close friend Dr. Kyle Stiegert was the main responsible person for this, as the foundational knowledge and the curiosity for much more acquired in classes and conversations with Dr. Stiegert – specifically in the way they he scaffolded and delivered them – are the primary reason I am able to “defend” this work today. Kyle “brought me” to Madison, giving me a new education, family, and home, and would be very proud to see me finally achieve the primary reason for him doing so. One of his most valued advice that will follow me forever was to not rename Narék when becoming an American citizen.

This work has also greatly benefited from comments and conversions with Drs. Daniella Molle, Edynn Sato, Grace Xin Li, Margo Gottlieb, Rebecca Callahan, Sara Kangas, Karen Thompson, Laurene Christensen, Eric Grodsky, Bradley Carl, Jed Richardson, Grant Sim, Tameka Porter, Michael Nicholson, David Kritzberg, Jeffrey Engels, Karla Wesley, Diep Ngyuen, Mariana Castro, Katherine Williams, Ken Bond, Rosalie Grant, Amber Wilke Fournier, Rachel Garett, and candidates Deonte Iverson, CJ Greer, and Gerald Dryer. Thank you for listening to me rattle on about English Learners, statistical models, and *Intersectionality*, and sharing some of your wisdom and ideas.

I thank WIDA's Research Subcommittee – a rotating group of dedicated educational leaders, professionals, researchers, educators, administrators, and practitioners representing WIDA's member State Educational Agencies. It has been an absolute privilege to "lead" this subcommittee for the last decade. Thank you for sponsoring and prioritizing this research inquiry by deeming it important for the students of the WIDA Consortium. James McCobb, Philip Olsen, Heather Villalobos Pavia, Barry Pedersen, Mary Nesbit-McBride, Nadja Trez, Robert Measel, Jan Lanier, Chandi Wagner, Wendy Perron, Marshall Foster, Mayra Valtierrez, Chris Williams, Laurel Williams, Veronica Salas, Adam Rollins, Jacob Barto, Hannah Gibble, Valen Cayetano, Andrea Degre, Raydel Hernandez, and many others have served on this committee and shared experiences from the states they represent. Thank you for your great ideas and collaboration on so many interesting and challenging research projects that have had and will continue to have a real impact on students' academic experiences and outcomes.

I thank many of my WIDA colleagues who have supported and listened to me present on this topic countless times over the last years. You have been very helpful and patient as I tried to navigate the challenges of pressing dissertation deadlines. Cathlin Foy, Alexis Glick, Katie Schultz, Elise Duskin, Sabrina Read, Elizabeth Cranley, Erika Rosales, Jason Kemp, Stephanie Gibson, Kim Varnell, Gordon West, Katie Stenz, Shea Head, Jon Nordmeyer, Anna Rhoad-Drogalis, Jonathan Gibson, Andrew Reichert, Fabiana MacMillan, Nick Kraninger, Rita MacDonald, Kris Stewart, Shannon DeWitt, Dan Machmeier, Fatima Rivera, Rhonda Stravinski, Mark Chapman, Rachele Stucker, Drake Akardi, Alicia Kim, Lynn Shafer Wilner, Jen Wilfrid, Drew Linkon, Samuel Aguirre, David MacGregor, Rebecca Quigley, Molly Waugh, Andree Miller, Anne Lies, Kyuongwon "Kei" Bishop, Elizabeth Warren, Meredith Trahan, Amy Kucharik, Annemarie Banas, Becky Holmes, Andrew Kahn, Syed Hadi, Miguel Angel Colon Ortiz, Karen Bach, Sandra Howard, Laura Marshall, Kimberly LeSage, Selena Franklin, Courtney Skare, Analleli Hernandez, Lodee Arnold, Terri Mosgrove, Heather Buettner, Ben Wheeler-Floyd, Hacer Kamarese, Delis Cuellár Klitkze, Luis Perea-Hernandez, Amada Spalter, Marcy Olson, Hannah Park, Matt Basler, Aaron Burreson, Kristin Graw, Vitaliy Shyyan, Lori Hanna, Xatli Stox, Jen Aleckson, Heather Elliott, Leslie Grimm, Hannah Hayes, Justine Kolb, Kelly Krahenbuhl, Jordan Gallant, Megan Frisbee, Alisa Rhoads, Sharon Pearce, Melissa Paton, Deedric Bauer, Veronica Machak, Jessica Davis, Fatima Qazi, Rachel Maske, Jorge Cardona, John Fischer, Lauren Bartholomae, Laura Wright, and many, many others have supported this work. Thank you, for your feedback and encouragements, making WIDA and ACCESS happen, the best professional, technical, and administrative support, and for your dangerously contagious Can-Do attitude! And a special thank you to the WIDA leadership team – Edynn, Gary, Grace, and Tim – for creating the time, space, and opportunities for this voluminous and extensive research project.

I am grateful to my parents for giving me all the opportunities one needs to achieve in academia. It is yet another privilege to have so many educators and teachers of different languages, mathematics, and sciences in my family; on the other hand, it also has been a lifelong, responsible, and redundant task, trying to prove to my English teacher mom that “my English is finally good enough.” My father would also be very proud, had I been on a faster trajectory towards my doctorate. I thank my brother for “holding the fort” with infinite patience, resilience, and grace. Thank you all for encouraging me to keep learning and exploring new horizons!

I thank my American “host” family, the Adams, who opened their door to a nerdy exchange student from the former Soviet Union. Your daily commitment to make sure that I experienced some of the great and dreamy things that America had to offer destined me to return to the Midwest to chase my own such dreams.

I am grateful to Michael Nicholson for his friendship and mentorship over the years. Thank you for the uncontracted, yet prudent investment in your Armenian translator’s graduate education. Alongside the invaluable insights gained in free-market economics and applied statistics through work on reform and development at USAID, as well as on the theory and practice of poker, this created an inevitable formula for success.

I am grateful to Jeffrey Engels for his friendship and mentorship over the years. In a selfless display of resistance that has since become an inspiration for me, you stood up against consular bureaucracy, disfunction, and red-tape, and enabled my dream to pursue my graduate education in the United States.

I am forever grateful to Mr. John Marshall Evans – the former ambassador of the United States to the Republic of Armenia. Mr. Evans was dismissed from his position in 2006 due to another such brazen and selfless display of resistance by uttering the G-word – only fifteen years ahead of the US official policy at the time – to do historical and moral justice and pay proper respect to the victims of the Armenian Genocide. Thank you for caring, knowing the price to be paid, but doing the right thing anyway!

I thank all my friends and relatives who have supported me throughout the years. There is no need to further call you out – you know who you are!

Despite the lack of any legible feedback on this project, I am grateful to Leo for his kind, calm, and measured friendship over the last decade. Thank you for the privilege of cohabiting with the perfect cat!

Finally, I am grateful to my partner, Anna. Thank you for your infinite kindness, wisdom, patience, support, care, and love. Thank you for sharing all the joys, hardships, and memories of the last quarter century. Many people listed above have uniquely contributed to me being here, in Madison, Wisconsin, USA, as I reach this milestone in the best school of education in the United States of America in 2024. But ultimately, you are the underlying reason and the best enabler!

Նարեկ

ABSTRACT

There is growing evidence that the COVID-19 pandemic had a large and negative impact on student learning. This dissertation explores the effect of the pandemic on one of the most vulnerable student subgroups: English Learners (ELs). In this work I examine existing disparities among subgroups of ELs, and the different ways the pandemic has impacted these disparities. These language learners' academic English proficiency, determining their status as an EL, is examined within the context of individual, institutional, and dynamic factors that have shaped and continue to impact these students' educational experiences and trajectories in American classrooms.

Leveraging population-level longitudinal data from ACCESS Online – an annual "high-stakes" language proficiency assessment administered across member states in the WIDA Consortium, I present evidence from regression models with increasing complexity that account for (a) the clustering of millions of students across thousands of schools, districts, and WIDA states, (b) individual-level factors such as students' time as EL, "newcomer" and "long-term" status, ethnicity and race, gender, disability status, interrupted education, migrant status, and parental refusal of language support services at school. Consistent and precise estimates from multilevel regression models highlight and document large disparities in the average English proficiency of ELs across several demographic subgroups, and provide timely and detailed data on the detrimental, differential, and ongoing impact of the COVID-19 pandemic on many young learners' academic outcomes. For example, students identified as Hispanic, making up most of the EL population, report substantially lower average proficiency compared to non-Hispanic identified ELs; the findings show that this disparity has further increased after COVID-19.

The uncovered disparities in proficiency between EL student subgroups representing multiple ethno-racial and other overlapping identities are interrogated under the theoretical lens of *Intersectionality* (Crenshaw, 1991), to identify, contextualize, quantify, and shed light on historical, political, and structural inequities in educational opportunities that result in systemic and persistent differences between academic outcomes. The *Intersectionality* framework, stemming from legal studies of Kimberlé Crenshaw, informs the complex and varying ways the pandemic has impacted English Learners' education, exacerbating the already-substantial disparities. The evidence shows some modest recovery for select EL subgroups; newcomer ELs and English learners across all racial identification categories report higher average scores than prior to the pandemic unless they also were identified with Hispanic ethnicity. Making up the majority of EL student population nationally, the findings emphasize the need for a more careful focus and more effectively designed support systems for these English language learners, who are consistently underserved and are falling further behind.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CSS	Composite Scale Score
EL/ELL	English learner / English Language Learner
ESL	English as a Second Language
ESSER	Elementary and Secondary School Emergency Relief
IEP	Individualized Educational Program
FEP	Fluent English Proficient
FE	Fixed effects
FTE	Full-Time Equivalency
GLS	Generalized Least Squares
LEA	Local Education Agency
LIEP	Language Instruction Educational Program
LTEL	Long-Term English Learner
NCES	National Center for Educational Statistics
OCR	Office for Civil Rights
OLS	Ordinary Least Squares
RE	Random effects
SEA	State Education Agency
SES	Socio-Economic Status
SLIFE	Students with limited or interrupted formal education

CHAPTER ONE: INTRODUCTION

In times of crisis, the most vulnerable are those who suffer the most. A recent quintessential example, the COVID-19 global pandemic wreaked havoc on every aspect of life, forcing the closure of schools for public health concerns in 2020-2021 and shifting to alternate modes of instruction to the extent local policies, circumstances, and resources allowed. States, districts, and schools struggled with continuing to provide quality education to students, and researchers predicted increasing inequities and disparities for many of the nation's underserved students. School buildings closed for safety concerns, and the rushed and chaotic switch to remote and hybrid modes of instruction highlighted and widened the educational disparities, as students from disadvantaged and marginalized backgrounds – who lacked access to technology, internet, and other resources enabling a conducive learning environment – faced even steeper obstacles in pursuing their education. Some years after schools reopened their doors to students in 2022, evidence is starting to emerge corroborating the early predictions that the disruptions to students' education caused by the COVID-19 pandemic had profound, complex, varied, and context-dependent impacts on students' lives and education.

The overarching purpose of this work is to build on, and extend this evidence, by: a) quantifying the cumulative impact of pandemic-induced negative shocks on the academic outcomes of English Learners – a student population that is often described as marginalized and underserved; b) identifying the more vulnerable subgroups within the very diverse EL student population through an analysis of disparities in outcomes measuring students' English proficiency; and c) assessing the impact of the pandemic on these disparities. The average and differential impact of the pandemic on EL outcomes and persistent disparities is estimated in the context of individual-, temporal-, and institutional-level factors, i.e., considering the student-level demographic data, the repeated nature of student assessment measures across time, and the enrollment/nesting of millions of ELs in schools, districts, and states.

I examine the impact of the COVID-19 pandemic on students' academic English proficiency in an interrupted time series framework, juxtaposing ELs' annually measured language proficiency before and after the pandemic. Leveraging large-scale population-level data from ACCESS for ELLs Online (hereinafter ACCESS) annual "high-stakes" language proficiency assessment used across states in the WIDA consortium (henceforth WIDA), the analytic strategy applies Ordinary Least Squares (OLS), longitudinal, and mixed-effects (hierarchical) regression models with increasing complexity that account for the clustering of the EL student population across thousands of schools, districts, and states, and quantify relationships between individual-level demographic factors and EL proficiency, thereby delineating important differences across multiple student categories. The empirical evidence gathered from multiple regression models highlights large disparities in the average English proficiency of ELs across several demographic subgroups, and provides timely and detailed data on the detrimental, differential, and ongoing impact of the COVID-19 pandemic on many young learners' academic outcomes.

More specifically, I estimate Ordinary Least Squares (OLS), longitudinal, and mixed-effects (hierarchical) regression models with increasing complexity that account for the clustering of millions of ELs (≈3.4 million unique ELs) across 43,183 schools, 7,619 districts, and 34 WIDA states, as well as for the individual-level factors such as students' estimated time as EL, "newcomer" and "long-term" status, ethnicity and race, gender, disability identification, interrupted education, and migrant status and waiver from supplementary language services at school. The ample empirical evidence gathered from these regression models confirms and documents large and persistent disparities in the average English proficiency of ELs across several demographic subgroups and provides timely and detailed data on the detrimental and ongoing impact of the COVID-19 pandemic on many young learners' disparate outcomes.

The uncovered disparities in proficiency between EL student subgroups representing multiple ethno-racial and other overlapping identities are interrogated under the theoretical lens of *Intersectionality* (Crenshaw, 1991) to shed light on historical, political, and structural inequities and disparities in educational opportunities that are reflected in systemic differences between academic outcomes. The *Intersectionality* framework, stemming from Black feminist legal studies of Kimberlé Crenshaw (1991), informs the complex and varying ways the pandemic has impacted English Learners' education, exacerbating the already substantial disparities, for example, between Hispanic and non-Hispanic identified EL students' average proficiency. Perhaps reflecting the recent efforts of the federal government to offset some of the pandemic-induced learning losses through Elementary and Secondary School Emergency Relief (ESSER) funding, the results also point to signs of some early post-COVID recovery for several EL subgroups. However, the findings also highlight the need for increased attention and better-targeted services for many more English Learners who may not be receiving the proper supports to develop high levels of academic English proficiency and exit EL status.

Terminology

The terms that permeate the landscape of English Learner education have been widely scrutinized (Brooks, 2017; 2018; Flores et al., 2015; Flores & Rosa, 2015; Kibler et al., 2018; Menken et al., 2012; Thompson, 2015). Acknowledging that some of these labels such as “English Learner” or “Long-term EL” are in essence deficit-based and can further stigmatize students who are still honing their multilingual skills, there has been a gradual and welcomed shift in the literature and in the field to a more assets-focused framing of these students, such as “multilingual learners” (MLs), “dual language learners” (DLs\DLLs), “emergent bilinguals”, “plurilingual learners”, and “ELs\DLs\MLs in extended stay”. While agreeing in spirit with such framing, in this work I use officially designated terms like *English Learners (ELs)* and *Long-term English Learners (LTELs)* to emphasize the definition and implementation of these student categories that is rooted in federal legislation, as well as its prevalent use in state and district regulations, rules, and policies. Further, this terminology is more aligned and appropriate for the analysis herein, since the examined sample includes exclusively those students who were, at the time their outcomes were measured, identified as English Learners as stipulated by federal law and implemented through state and district regulations and rules where they were enrolled.

Finally, and perhaps most importantly, while there have been some positive shifts in the discourses towards asset-based descriptors of this student subgroup, as the evidence presented by this study confirms, the academic outcomes that are reported by the students assigned to these labels cannot claim similar improvements. Therefore, more needs to be done, in addition to using better terminology, to ensure that these young language learners are provided with educational opportunities that all children deserve.

Positionality as a Researcher

Alike many of the English learner students taking the ACCESS assessment who are racialized or ethnized into discrete identity categories, when asked to reveal my ethno-racial identities I ascribe to the “non-Hispanic” and “white” categories provided in surveys and forms while I want to represent so much more than these simplistic checkboxes allow. (South) *Caucasian* by geographic birthplace, white as measured by *albedo*,[1] and *White* as “ordained” by the US Supreme Court in 1924,[2] my additionally overlapping Armenian identity also ascribes to historical roots steeped in ethnic, linguistic, religious, economic, and political oppression and assimilation throughout centuries, as forced by Assyrian, Byzantine, Roman, Arab, Persian, Mongol, Tatar, Seljuk, Ottoman, and most recently – Soviet empires, the hold of which collapsed when I was ten years old. A very small part of Armenia has managed to prevail as an independent state with its millennia-old language, religion, and culture. However, the outlook of Armenia remains very bleak. At the troubled crossroads of major geo-political highways (dis)connecting Russia and Iran on the North-South, and Azerbaijan and Turkey on the East-West directions, Armenia, for me, is the quintessential example of how dynamic, multiply overlapping, and conflicting *Intersectional* forces can shape, marginalize, deprecate human and lives. Moreover, I believe that by holding on to over a century-old memory of

---

[1] Albedo is a scientific term, measured to capture the fraction of sunlight that is diffusely reflected by a body.

[2] In “On the Boundary of White: The Cartozian Naturalization Case and the Armenians, 1923-1925,” Craver (2009) highlights the advantaged socio-economic position of the plaintiff, and writes: “… the survival of the 1923 naturalization challenge in U.S. v. Cartozian helped ease the Armenians' way in American society, whereas their experience would not be comparable to that of the African Americans or the Asians”, p.51.

Genocide [3] – the ultimate form of racism – and persistent drive for its (official) recognition, many Armenians carry a generational trauma, but also an additional responsibility to recognize, call out, and oppose biases leading to discriminatory systems and inequitable outcomes.

Having lived half of my life in Armenia and the other half in the American Midwest, (roughly two decades each) my additionally overlapping immigrant identity also informs and motivates this study. In this work I examine EL proficiency, a construct that represents both these students' "language ability status", and a gateway to educational and career opportunities for many immigrant students. Therefore, it is of utmost importance to ensure that all young language learners have access to equitable educational opportunities, as graciously promised and legislated by the US government.

I also recognize that my life experiences are inseparable from my outlook on the topics of systemic oppression, discrimination, and racism, and its educational consequences that I explore in this work. My perspectives as a researcher are impacted by the biases and assumptions I may hold because of my negative and positive[4] experiences shaped by these (and other) intersecting identities. Despite its scale and scope, the student assessment data that underlies the analysis in this work could not "speak for itself", for example, with respect to the uncovered disparities in academic

---

[3] After decades or denial, neglect, and relabeling, the US official policy changed in 2021 on this issue. On April 24, 2021, Armenian Genocide Remembrance Day, President Joe Biden declared that the United States considers these events "genocide" in a statement released by the White House. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/24/statement-by-president-joe-biden-on-armenian-remembrance-day/ . Retrieved 2024-04-24.

[4] An example of such a "positive experience" is my identity as a multilingual individual, due to Armenia's location in the intersection of ex-Soviet and Western forces when I was growing up, driving me to learn Russian and English, in addition to my native Armenian. Working as an interpreter and translator for international organizations and Western development missions in post-Soviet Armenia amplified my exposure to (western) intersectional forces and further drove me to pursue graduate education in the United States.

outcomes without the adopted *Intersectional* lens informing and motivating critical research questions and guiding the analytic strategy. Moreover, as demonstrated in the findings, the adopted *Intersectional* lens provides a sharper focus and additional nuance to the data and findings and allows me to speak louder and clearer about these disparities.

Organization of Chapters

This work is organized as follows. Chapter 1 started with an Introduction, outlining the essential components of the study. Chapter 2 provides a brief review of the literature, focusing on the recent work discussing current issues in English Learners and Hispanic students' education, including on the predicted and apparent detrimental impact of the COVID-19 pandemic. I review research that highlights disparities within the very diverse EL population and discusses how overlapping individual-level disadvantages accumulate and can lead to increasingly unfavorable outcomes. Next, I provide a brief overview of the *Intersectionality* framework, its main principles and methods of analysis, and few applications in examining English Learners' education. Building on existing research, I present a *Framework of English Learner Intersectionality*, with English Learner students' status as EL centered around overlapping and intersecting student identities and nested in multileveled systems of education.

Chapter 3 on Methods begins with a presentation of the research questions and the conceptual model applied to interrogate the differential impact of the COVID-19 pandemic on EL proficiency and status, considering the complex and intertwined nature of intersectional identities, and its relationship with temporal- and institutional-level factors. Next, I present the underlying data, including a description of the variables and descriptive statistics on students' demographic and educational outcome data. The chapter concludes by outlining the proposed analytic strategy of using regression as a method of analysis, its connections to the theoretical framework of *Intersectionality* and describes the multiple primary and auxiliary regression models applied to interrogate ELs' proficiency outcomes.

Findings, presented in Chapter 4, describe the main model parameters estimated by various regression models with increasing flexibility and complexity, and auxiliary model parameters estimated for extensions of these models aimed at: (a) providing meaningful comparisons of statistical relationships across multiple levels of analyses; and (b) ensuring robustness and consistency across various model specifications. More specifically, I quantify the average impact of COVID-19 on EL proficiency using OLS, longitudinal, and mixed-effects models, and outline the uncovered disparities in EL proficiency outcomes. Next, I demonstrate the differential impact of the pandemic on student outcomes, by outlining how the disparities and differences across focal student subgroups have changed in the aftermath of the COVID-19 pandemic.

Chapter 5 concludes this work by providing a discussion of the results from the empirical analyses. I summarize the main findings of the study and provide some potential implications for research and practice. I outline the number of ways this work contributes to the existing literature, list several caveats and limitations, and present suggestions on how future research and analyses can further inform the inquiry. Included Appendices A and B provide details on the empirical results not included in the main presentation.

CHAPTER TWO: LITERATURE REVIEW

Introduction

Due to the wide range of variables impacting English Learners and academic outcomes that are of research interest, and the vast scale and scope of the emergent literature on ELs' education that highlights persistent gaps in the educational opportunities and outcomes of these students compared to their never-EL peers, an extensive literature review is not feasible in the scope of this work. A few, but much more rigorous accounts on the multitude of challenges and important questions around English Learners' education are available, such as the comprehensive review titled "Promoting the Educational Success of Children and Youth Learning English" by the National Academies of Sciences, Engineering, and Medicine published in 2017. There is also a large body of work on EL education with more specific foci, ranging from general immigration and education policy (Callahan et al., 2023; Gándara & Rumberger, 2009; Umansky & Porter, 2020; Sugarman, 2019; Villegas & Pompa, 2020), to appropriate and accurate identification of ELs (Abedi, 2014; Artiles & Ortiz, 2002; Bailey & Kelly, 2013; Cook & Linquant, 2015; Lopez, et al., 2016), to adequate, effective, and equitable academic language and content instruction (Bailey & Heritage, 2014; Calderon et al, 2011; Cummins, 2021b; DiCerbo et al., 2014; Menken & Kleyn, 2010; Molle et at., 2015 Stephens & Francis, 2018), and timely exit / reclassification from EL status (Cimpian et al., 2017; Linquanti & Cook 2015; Kieffer & Parker, 2016; Robinson-Cimpian & Thompson, 2016; Schissel & Kangas, 2018; Umansky & Reardon, 2014), among many others. The literature on the education of long-term ELs (Brooks, 2018; Clark-Gareca et al., 2017; Kim & Garcia, 2014; Olsen, 2014 & 2010; Sahakyan & Ryan, 2018; Shin, 2020;

Umansky & Avelar, 2023; Villegas, 2023), dually-identified (ELs with disabilities) students (Akerman & Tazi, 2015; Burr et al., 2015; Buenrostro & Maxwell-Jolly, 2021; Hamayan et al., 2013; Kangas, 2014; 2017; 2018; Murphy & Johnson, 2023; Sahakyan & Poole, 2022; Shifrer et al., 2011), ELs from immigrant, refugee, and otherwise-interrupted education backgrounds (Callahan et al., under review; Callahan & Humphries, 2016; Darling-Hammond 2010; Gándara & Contreras, 2009; Hopkins et al., 2015; US Department of Education, 2016) is also vast, interconnected, and reflective of the complexity and plethora of issues that permeate the education of EL students.

The approach I take in this chapter is to provide a brief review of the recent research on most current issues in English Learners' education, with a focus on recent quantitative studies that have started to predict and present evidence on the impact of COVID-19 on ELs' educational opportunities and academic outcomes. I summarize recent research that interrogates issues around inequities in English Learners' and Hispanic students' education, and those that highlight variations and differences in the ways multilingual students' education is organized and implemented across schools, districts, and states. Studies that examine English Learner outcomes while directly considering the tremendous diversity of this student population and provide evidence of differential performance across subgroups are highlighted. Finally, I provide a brief review of the adopted theoretical framework of *Intersectionality* (Crenshaw, 1991), with a focus on its few applications to inform ELs' education and ways in which individual, temporal, and institutional factors interact with students' multiple and overlapping individual identities.

English Learners and Disparities in Outcomes

In 2020, about 10% of K-12 students were identified as English Learners (ELs) across the United States (NCES, 2023). ELs receive secondary language support services until they meet state-established criteria for reclassification. English Learners are one of the fastest growing student populations, as some estimates project that by 2025 one in four students in US schools will be identified as an EL (NEA, 2020).[5] English Learners are also one of the most vulnerable and marginalized student subgroups; large and consistent disparities in educational opportunities and subsequent academic outcomes between EL and their monolingual English-speaking peers have been extensively documented and discussed in the literature (Callahan & Shifrer, 2016; Fry, 2007; Johnson, 2022; Kieffer & Parker, 2017; Kanno & Kangas 2014; NCES, 2019; Ream et at., 2017). ELs are more likely to drop out of school (Boone, 2013; Callahan, 2013) and less likely to complete high school and attend college than their non-EL counterparts (Callahan et al., 2010; Kanno & Cromley, 2013; Ruiz-de-Velasco & Fix, 2000). These persistent disparities in opportunities and achievement have historical roots in discrimination against non-native English speakers (Bonilla-Silva, 1996 & 2015; Russell, 2023) and have been attributed to higher rates of poverty, higher mobility rates, and the greater likelihood for ELs to attend segregated, underfunded, and unsafe schools, compared with their non-EL counterparts (Darling-Hammond, 2007; Fry, 2008; Garver, 2020; Jiménez-Castellanos & García, 2017; NCES, 2019; Olivares, 2022; Rodriguez, 2020; Sahakyan & Cook, 2014). In a seminal reframing of this discourse, Ladson-Billings (2006) suggests considering “education debt” rather than “academic gaps” and

---

[5] https://www.nea.org/resource-library/english-language-learners

differences in outcomes. Milner (2012) further argues that opportunity gaps are more comprehensive than just differences in standardized test scores and are shaped by interconnected social, economic, and political factors. While other critics of such gap- and deficit-focused analyses point out that standardized tests that are administered in English may obscure and underreport the true knowledge and abilities of young learners who are still developing academic English (Faulkner-Bond & Sireci, 2015; Saunders & Marcetelli, 2013; Hopkins et al., 2013; Ream et at., 2017; 2017), other researchers further report that many of these disparities are likely driven by factors related to students', their families', and schools' socio-economic status (Adair, 2015; Butler & Le, 2018; Carhill et al., 2008; NASEM, 2017; Kieffer, 2010; Kim et al., 2014). For example, according to the Migration Policy Institute's analysis of US Census data, one-third of the immigrant population, or about 15 million people, were "low-income", reporting levels 200% below the federal poverty line. Moreover, the report found that approximately two-thirds of these low-income immigrants identified as Hispanic (Gelatt et al., 2022).[6] And while a substantial number of ELs are born in the US (García & Kleifgen, 2018) research shows that English Learners are predominantly immigrant-origin, i.e., the children of foreign-born parents (Callahan & Humphries, 2016).

Further, perhaps due to lack of reliable large-scale data across contexts, apart from a few notable exceptions (Callahan et al., 2010; Dorn et al., 2020; NASEM, 2017; Slama, 2014; Umansky et al., 2020), there has been little research on ethno-racial, or other disparities in outcomes within the incredibly diverse EL population, and especially how the pandemic may have affected these disparities. English Learners have complex

[6] The next highest proportion identified as Asian and Pacific Islander, estimated at about 20%. Retrieved from: https://www.migrationpolicy.org/sites/default/files/publications/mpi_low-income-immigrants-factsheet_final.pdf

and intersecting identities besides their EL status that should be highlighted and considered. Indeed, among commonly cited weaknesses of several quantitative studies examining the impacts of the pandemic on student outcomes are the non-diverse samples and small sample sizes underlying the analyses (Garbe et al., 2020; Martinez & Broemmel, 2020; Marshall et al., 2020; Schaefer et al., 2020).

Disparities for Hispanic Students

Hispanic students make up the majority of the EL population nationally, and about 70% of the student population in WIDA states. Studies show that they have experienced unequal access to school funding, high-quality teachers, educational materials, instructional time, course offerings, and adequate facilities (Baker, et al., 2020; Carnoy & Garcia, 2017; Gándara & Rumberger 2009; Gándara & Orfield, 2012; Orfield et al., 2016; Rumberger & Gándara 2004).[7] Reflecting these inequities, Slama (2014) reports that it takes almost twice as long for Hispanic ELs to attain English proficiency compared to their non-Hispanic peers. Other studies also corroborate that EL reclassification rates are typically slower among Spanish-speaking students and those from disadvantaged socio-economic backgrounds (Conger, et al., 2009; Thompson, 2012; Kao & Thompson, 2003; Kim et al., 2015; Umansky & Reardon, 2014). This is important considering that students who reach reclassification-level proficiency before middle school tend to outperform their never-EL peers on standardized assessments (Hill et al., 2014; Saunders and Marcetelli, 2013), whereas students who remain in EL status for prolonged periods of time are prevented from enrolling in advanced coursework (Callahan, et al., 2010; Lillie et al.,

[7] Consistent with the terminology used in the students' assessment recording the test takers' ethnicity, among other demographic data, I use the term Hispanic, instead of Latino/Latina/LatinX/Latiné, unless expressly used by the literature source.

2012; Umansky, 2014). Indeed, the negative effects of racialized and exclusionary tracking for ELs and Hispanic students have been well documented by research (Callahan & Muller, 2013; Callahan & Shifrer, 2016; Estrada & Wang, 2018; Gamoran, 2010; Kangas & Cook, 2020; Umansky, 2016). Moreover, Harklau (2016) provides evidence that the educational system may be especially misaligned with the needs of these students and explains that “the bureaucratic nature of schooling and a constant onslaught of bureaucratic errors and omissions is partly responsible for the high school underachievement in Hispanic children or immigrants.” p.601.

There is ample and growing evidence that Hispanic students and ELs face large and persistent disparities in educational opportunities and achievement as compared to other students. As outlined in the next section, there are many concerns, predictions, and growing evidence that the COVID-19 pandemic has exacerbated many of the existing academic disparities for these students.

English Learners and COVID-19

The COVID-19 pandemic exposed vulnerabilities in social and economic systems that chronically underinvest in essential public services (United Nations Human Rights Office, 2022). The pandemic also had a profound impact on the K-12 education system nationally, with schools forced to close their doors or adopt remote and hybrid learning approaches. However, disruptions to in-person instruction and shifts to hybrid or virtual classrooms have affected different subgroups of students in diverse ways. Evidence is already mounting that English Learners have been among the students hardest hit by COVID-19’s disruptions to in-person learning (OCR, 2021; Nowicki, 2020; Huck & Zhang, 2021) and that ELs faced significant challenges during and after the pandemic due to many systemic factors and individual circumstances that increased their vulnerability (Santibañez & Guarino, 2021). [8, 9] Along with access to teachers trained in language development and modified course content, ELs require carefully calibrated and scheduled, intentionally scaffolded, and appropriately delivered school- and program-based supports (August & Shanahan, 2006; Boals, et al., 2015; Daniels & Westerlund, 2018; NASEM, 2017; Nordmeyer et al., 2021; Rumberger & Gándara, 2000). Therefore, disruptions in these and many other critical elements of ELs' education brought about by the pandemic likely exacerbated existing inequalities (Bacher-Hicks et al., 2020; Dorn et al., 2020; Hamilton et al., 2020; Villegas & Garicia, 2021). Even prior to the pandemic researchers had voiced concerns over the "digital divide", with Black, Indigenous, and other students of color having restricted access to technology and high-speed internet (Alliance for Excellent Education, 2020; Education Trust, 2020), and suggested that online schooling can come with an online penalty for struggling and vulnerable learners (Dynarski, 2018; Zehler et al., 2019) and for Hispanic learners (Kaupp, 2012) and that factors like technological support at home and in school, as well as prior high achievement and self-discipline were essential for an effective online learning experience (Heissel, 2016; Villegas & Pompa; 2020). Even when ELs had access to the technology necessary for online learning, research shows that remote and hybrid modes of instruction typically limit social and peer-to-peer exchanges due to a truncated instructional model and

---

[8] Department of Education, 2021. Office for Civil Rights. Education in a Pandemic: The Disparate Impacts of COVID-19 on America's Students. Retrieved from: https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.pdf

[9] US Government Accountability Office, 2020. Distance Learning: Challenges Providing Services to K-12 English Learners and Students with Disabilities during COVID-19. Retrieved from: https://www.gao.gov/products/gao-21-43

reduced opportunities of low-stakes interactions (or informal conversations between peers in the cafeteria, during recess, and between classes) and collaboration-focused groupwork, which are important components of English Learners' basic language proficiency development (Baruch, 2023, Echevarria et al., 2017, Molle & Lee, 2015).

Further, as schools across the country struggled with organizing safe in-person instruction due to staff shortages (Bryner, 2021; Mason-Williams, 2020; Rosenberg & Anderson, 2021), limited English comprehension and longer working hours deterred many EL families from assisting their students with the digital technology and modified curriculum at home (Nowicki, 2020). Further, students learning English are often tasked with additional, out-of-school activities, such as caring for siblings and serving as translators and interpreters to help struggling adults (Huck & Zhang, 2021; Rodriguez et al., 2020). These and other significant "pull-out" factors were exacerbated by the additional challenges brought about by the pandemic.

Considering that EL students are engaging with challenging academic content while mastering an additional language (Calderon et al., 2011; Clark-Gareca et al, 2020; Cook, et al., 2011; Solórzano, 2008) falling behind in English language acquisition in early stages of their academic development can have a cumulative and detrimental impact on these students' educational and career trajectories (Sugarman & Lazarin, 2020; Tindal & Anderson, 2019; Stevens & Schulte, 2017). As summarized by Sugarman and Lazarin (2020) in a Migration Policy Institute brief published five months into the pandemic to provide guidance to states and districts on immediate strategies aimed at preventing learning loss for ELs, *"…despite these long-standing legal protections to ensure equitable access to education, the pandemic has shined a spotlight on how tenuous such policies* *are in many parts of the country. And despite heroic efforts on the part of many educators to provide their students access to instruction during school building closures, existing weaknesses within the school system … have rendered such efforts ineffective" (p.3).*[10]

Given the importance and urgency of the issue for the large and growing population of English Learners, researchers have already started assessing the impact of the pandemic on English Learners' education. For example, Baruch (2023) examined the impact of remote learning on Delaware's EL students using ACCESS data from 2016 to 2022 and reported a large and significant negative English proficiency growth rate during the period of virtual learning, followed by relatively weak growth during the final year. Other reports, such as Sahakyan and Cook (2021), Sahakyan and Poole (2022), and Poole and Sahakyan (2023) provide descriptive evidence of trends on the proficiency and annual growth of English Learners across the WIDA Consortium, and report substantial differences in pre- vs post-pandemic scores of EL students in all grades and most individual language domains. Further, Poole and Sahakyan (2024) provide corroborating evidence of a large and ongoing impact of the pandemic on student scores based on the most recent, 2023 school year data. Importantly, Poole and Sahakyan (2024) present descriptive evidence of existing disparities between the proficiency outcomes of ELs identified as Hispanic compared to non-Hispanic English learners, which have seen an increased following the COVID-19 pandemic. However, while these reports leverage very large samples of EL student outcomes, the authors invite caution in interpreting the results due to the many context-dependent individual- and institutional-

[10] Educating English Leaners during the COVID-19 Pandemic: Policy Ideas for States and School Districts. September, 2020. Retrieved from: https://www.migrationpolicy.org/sites/default/files/publications/mpi-english-learners-covid-19-final.pdf

level factors that might mask important differences in the outcomes of this very diverse student population.

Apart from these large-scale (albeit descriptive) studies, most of the recent research that examines the impact of the pandemic on EL language development are cross-sectional, based on small, selected, or non-diverse samples, depend heavily on the local context, and typically include data on students from a single state, district, school, or cohort (Johnson, 2023; Huck & Zhang 2021). Moreover, while generating useful and much needed evidence on English Learners' academic outcomes, these analyses overlook the important role of nested institutional relationships, as reflected by state, district, and school-level hierarchies. For example, starting with the Bilingual Education Act (1968) and the Lau v. Nichols (1974) Supreme Court decision, which mandated that schools must provide a meaningful education to EL students, the interpretation of federally-defined laws, regulations, and policies is undertaken at the state-level by State Education Agencies (SEAs), followed by their further unpacking and implementation at the district-level by Local Education Agency (LEA) administrators and educational officers, and finally at the school-level, by principals, school administrators, staff, and teachers. These hierarchies and nested processes could result in a diverse range of practices in how states and districts cater to the needs of EL students (Bartlett et al., 2024; Bond, 2020; Mavrogordato et al., 2022; Callahan et al., 2021; Callahan & Hopkins, 2017; Villegas & Pompa, 2020).

In sum, as reported by Huck and Zhang (2021) in their review of the early literature on the impacts of COVID-19 on the general K-12 education landscape, "student outcome data is needed to support predictions of learning loss and the extent to which achievement *gaps* have widened*"* (p. 73; emphasis added, see footnote).[11]

Answering the call, taking into account individual-level demographic characteristics and the institutional context of state-, district- and school-level factors, this study provides timely, consistent, comprehensive, and generalizable empirical evidence on a) the overall, average "learning losses" the English Learner population has incurred throughout and after the pandemic, and b) the extent to which various subgroups of EL students have been impacted in their learning trajectories and thus need more immediate attention and support.

In addition to the studies reviewed in this section, three seminal books have guided the selection, implementation, and interpretation of theoretical and empirical methods and models. Michael Russells' "Systemic Racism and Educational Measurement: Confronting Injustice in Testing, Assessment, and Beyond" (2023), Cornell & Hartmann's "Ethnicity and Race: Making Identities in a Changing World (Sociology for a New Century Series, 2007)", and Rabe-Hesketh and Skrondal's "Multilevel and Longitudinal Modeling Using Stata Volume I: Continuous Responses" (2021), have all been essential in informing the various components of this study. The theoretical framework, reflected in all of these components, is largely based on the scholarship of Kimberlé Crenshaw on *Intersectionality,* and is presented in the next section.

[11] In this work I use the term "disparity" instead of more neutral sounding "gap", to emphasize that an intentional disturbance in parity exists which is calling for further action.

Theoretical Framework

The theoretical framework I adopt to interrogate disparities in educational outcomes within the EL student population, and the impact of COVID-19 on these disparities, is grounded in *Intersectionality* – a body of work that explores the compounding and marginalizing effects from multiple categories of minoritized student backgrounds (Crenshaw, 2013 & 1991; Hankivsky & Cormier, 2019; Schissel & Kangas, 2018; Russell, 2023). *Intersectionality*-focused approaches emphasize the importance of examining, understanding, and challenging the ways systematic racism shapes the work of institutions, and caution about policies and practices that may appear neutral or benign, but further harm historically marginalized groups (Crenshaw, 2011; Delgado, 1995; Ladson-Billings & Tate, 1995). Originally stemming from Black feminist legal studies of Kimberlé Crenshaw (1991) that examined the failure of antidiscrimination laws to address Black women's distinctive, intertwined experiences of racism and sexism (Wang, 2023), the framework of *Intersectionality* examines the relationships between overlapping social identities (e.g., based on gender, race, ethnicity, ability, etc.) and the related structures that create and perpetuate systems of oppression. As social identities intersect at the individual level (e.g., race and gender), experiences at those intersections are influenced by larger interpersonal and structural systems of oppression such as racism and sexism (Bowleg, 2012; Collins, 1995).

According to Hankivsky (2014), the central tenets of *Intersectionality* assert that:

- *human lives cannot be reduced to single characteristics;*
- *human experiences cannot be accurately understood by prioritizing any one single factor or constellation of factors;*

- *social categories/locations, such as race/ethnicity, gender, class, sexuality and ability, are socially constructed, and dynamic;*
- *social locations are inseparable and shaped by interacting and mutually constituting social processes and structures, which, in turn, are shaped by power and influenced by both time and place; and,*
- *the promotion of social justice and equity are paramount.*

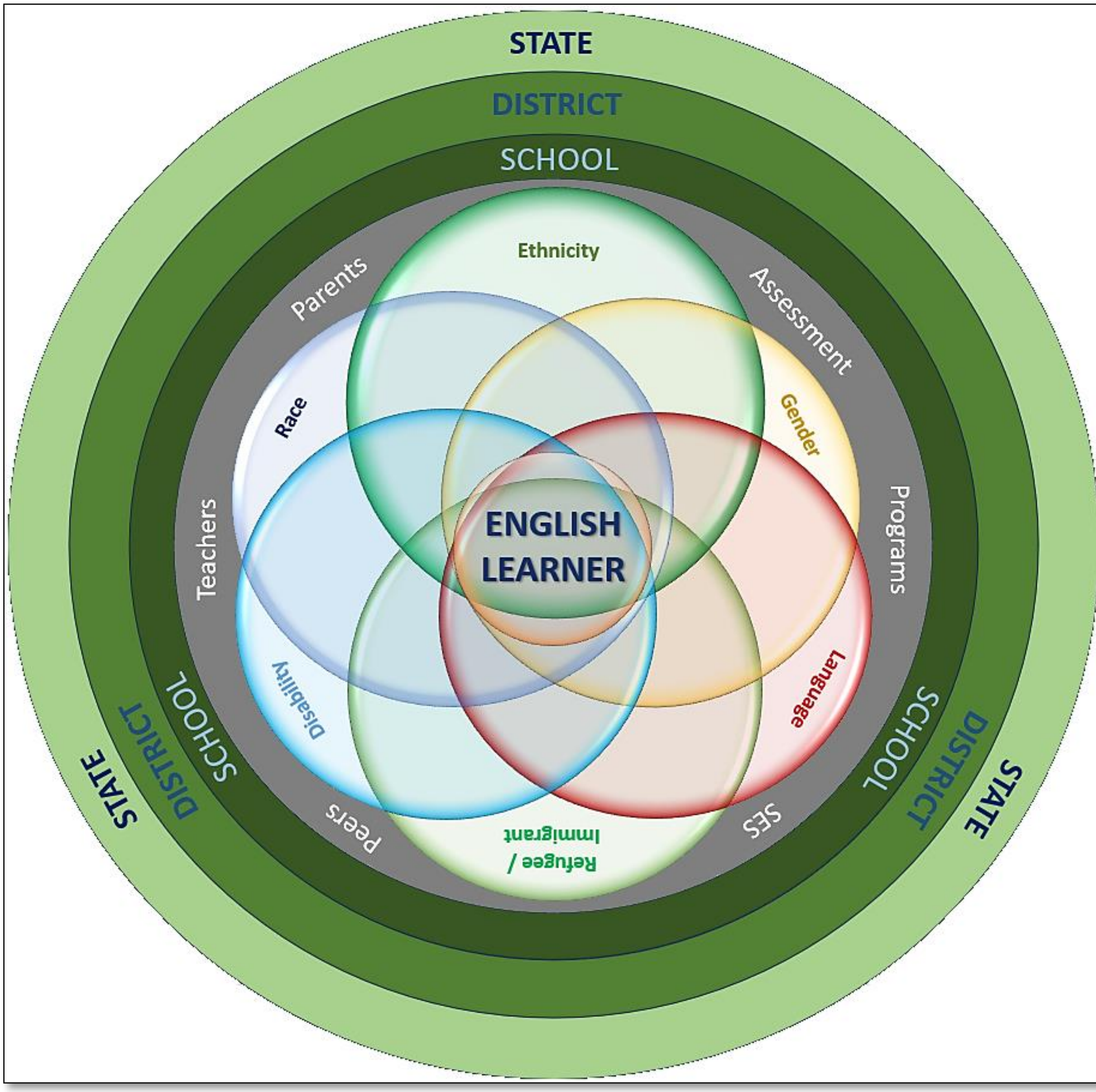
Echoing these principles and building on the work of Jiménez-Castellanos and García (2017) who conceptualize the "multiple lived realities of an English Language Learner" in a mosaic (Figure 1, p. 436), the theoretical framework in Figure 2.1 represents students' English language proficiency, driving students' subsequent status as an English Learner (for at least another academic year) as another socially constructed category that is nested within structures, policies, and practices at all the levels of the education system.[12] Students' status as an English Learner is therefore centered in a vortex of institutional [13], i.e., state-, district-, and school-level factors that interact with ELs' multiply-overlapping social identities in different and dynamic ways.

Following other researchers who focus on lines of inquiry aimed at disrupting political and structural inequities in educational opportunities (Artiles, 2013; Bonilla-Silva, 1996; Cho et al., 2013; Hankivsky et al., 2019; Kanno & Kangas, 2024; Sahakyan & Poole, 2023; Schissel & Kangas, 2018) English Learner students' academic outcomes and disparities in this analysis are examined, contrasted, and interpreted within the context of an educational system that has a profound effect on students by assigning

[12] In addition to the three core components of ethnicity, language and class, and race, Jiménez-Castellanos and García (2017) add intersections with religion, gender, race, and immigration status (figure 1, p. 436).
[13] Welton et al., (2018) underscore the appropriateness of the term "institution" as historical unspoken norms and social agreements become "instituted" or developed over time.

membership to subgroups through various state-, district- and school-specific EL identification, instruction, and reclassification policies.

Figure 2.1. Theoretical Framework: English Learners' Intersectionality.



Following the tenet of *Intersectionality* highlighting the importance of striving for social justice and equity, it is only natural to demand that these systems be held accountable for and better address the increasingly disparate outcomes that are subsequently demonstrated by many marginalized and vulnerable students.

Further, the tenets of Intersectionality are closely aligned with the guiding principles of *Quantitative Critical Race Theory (*QuantCrit*),* which is a framework that integrates the critical examination of race and inequality with quantitative methodologies (Garcia et al., 2017; Castillo & Gillborn, 2023; Tabron & Thomas, 2023). While both *Intersectionality* theory and QuantCrit involve examining how power structures can further marginalize vulnerable communities, and insist on the non-neutrality of the data and the socially-constructed nature of (ethnic and racial) categories, QuantCrit situates race, and as in the application of the present study – (Hispanic) ethnicity – at the center of discussion, making it explicit that findings must be interpreted within the context of historical, economic, and structural inequalities (Bonilla-Silva, 2015; Zuberi, 2001).

Consistent with prior research that has documented considerable variability across states, districts, and even schools in which EL students are identified, educated, and reclassified (Cimpian, et al., 2017; Linquanti & Cook, 2015; Kieffer & Parker, 2016; Kim et al., 2018; Estrada & Wang, 2018; NASEM, 2017; Villegas & Pompa; 2020), this work contributes to the literature by providing evidence that many student subgroups that share the overarching "English Learner" designation report consistently disparate educational outcomes, and that institutional contexts matter in how these disparities are shaped and affected. The included variables reflecting students' reported identities and their intersections are some examples of individual-level factors that can help surface ways in which inequitable educational opportunities, and subsequently disparate academic outcomes, are manifested for many English learners.

Moreover, in addition to the individual identities shown in Figure 2.1 (and others that are not listed), there are other important factors and circumstances that can impact EL students' proficiency. Depicted in the grey zone in white font, some examples of such factors are the types and/or the quality of the language support programs (LIEP) EL students are enrolled in, or the impacts of other in- and out-of-school supports for academic learning as provided by ELs' teachers, peers, and parents, all in turn potentially moderated by the students' and families' socio-economic status (Kao and Thompson, 2003; Le et al., 2024; Mavrogordato & White, 2017; Schmid, 2001*)*. These various impacts and effects interact at varying degrees with individual-level and institutional-level factors and shape English learners' outcomes, as typically measured through various standardized assessments which are administered in English, and in turn have been criticized for not capturing students' true abilities (Acosta et al., 2020; Faulker-Bond & Sireci, 2015; Solórzano, 2008).

Differently colored overlapping circles in Figure 2.1 are intended to highlight some of the intersecting identities that are pertinent to the education of ELs at the individual-level, and that this analysis explores. Not labeled in Figure 2.1 are the many overlapping regions that are multiply-highlighted and shaded by the "higher-level" identities. For example, students identified as ELs (at the center of the graph) can also simultaneously be located at intersections of various races and Hispanic ethnicity, or at the intersection of ethnicity and gender, or ethnicity and disability, both also considered herein. While this dissertation explores some of these overlaps, as detailed in the conceptual model in the next chapter, it is beyond the scope of any single work to investigate the many potentially relevant doubly-, triply, and multiply-intersecting identities.[14]

Further, Figure 2.1 presents a simplified view of the overlapping intersectional identities, in that these are depicted as symmetrical and appear very proportional. The interactions between overlapping identities and their intersections, and institutional-level factors are uniquely different for each of WIDA's three and a half million students that were captured by the overarching English Learner category and included in the empirical analyses. Each EL's collection of circumstances can be viewed as a unique, kaleidoscopic configuration of many of these multileveled, multifaceted, and intertwined factors.

Also not captured in this conceptual, two-dimensional *Framework* of *EL Intersectionality* is the ever-important element of time, rendering the multileveled relationships and interactions presented in Figure 2.1 into a complex, yet simplified snapshot, centered around ELs' cumulative English proficiency outcome at the time of the annual assessment of proficiency. Meanwhile, the *Intersectionality* lens invites a special focus on the fluid and dynamic nature of the factors and processes affecting students' academic outcomes. Indeed, EL proficiency and status – as shaped by overlapping individual identities, and manifested though students' interactions and relationships with institutional-level factors – is also continuously impacted by dynamic changes such as a global pandemic that can transform individual circumstances, and institutional policies, and educational resources in drastic and differential ways.

[14] Also not shown in Figure 2.1 is the federal level, which impacts the entire educational system (of ELs) profoundly. The tremendous variation in the implementation of different types of federal legislation and policies occurs at the state and lower levels, which are shown in Figure 2 and included in the empirical analyses.

The intertwined and context-dependent nature of the many multileveled factors, only some of which are observable, or perhaps measurable only to a certain extent, complicates rigorous large-scale quantitative inquiries into student outcomes. Nevertheless, identifying and highlighting disparities in students' proficiency attainment over time that may be attributable to factors that we do observe and measure can provide important insights into the types of systematic inequities that – regardless of the potentially omitted data – exist, persist, and have increased, especially for more marginalized EL student subgroups. Leveraging the large-scale data available for examining many of the multileveled factors, the purpose of this study, therefore, is to evaluate and quantify differences in outcomes that can be linked to such factors.

The theoretical framework of *Intersectionality* informs and gives perspective to several core components of this study. Prioritization and organization of research questions on examining existing disparities across many EL subgroups and impacts of the pandemic on thereof, decisions impacting the analytic strategy, and selection of specific regression methods, the scope and scale of included data, and the focal variables of interest that identify intersectional student groups of research interest - have all been informed and given perspective by the guiding principles of *Intersectionality*. These perspectives and connections are further highlighted in the relevant sections describing the data, methods, analytic strategy, and findings. To make these perspectives and connections more explicit, I italicize the term *Intersectionality* throughout this work.

CHAPTER THREE: METHODS

Introduction and Organization

In this chapter, I present the research design, data sources and underlying analytic sample, and analytic strategy applied to examine relationships between English Learner students' average proficiency, various individual- and aggregate-level factors affecting it, and the impact of the COVID-19 pandemic on this proficiency.

In the first section, I describe the research questions motivating the study, and the conceptual model that operationalizes these questions. Next, I offer an overview of the data, and discuss the key variables of research interest. Presenting the hypothesized problem statement – i.e., the potentially large and differential impact of the pandemic on English Learners' proficiency – in data terms, I provide a descriptive account of the observed aggregate trends of English proficiency across time, and for focal student subgroups.

Next, I describe the rationale for applying regression methods to interrogate EL proficiency outcomes in the context of the COVID-19 pandemic and through the lens of *Intersectionality*. Selection of the specific covariates and the regression models for quantifying relationships and decomposing variations in ELs' measured proficiency by individual, institutional, and temporal factors is further explained. The section concludes with a description of some of the properties and features of the specified regression models. To aid readers in navigating between the large number of specified models and presented variables, in these sections I capitalize and italicize the text when specifically referring to variable or model names.

Research Questions

Given perspective through the *Intersectionality* lens and supported by the entire set of WIDA (online) assessment data, this study aims to identify and quantify potentially differential impacts of the COVID-19 pandemic on English Learners' educational outcomes, while considering the multiple overlapping intersections of students' EL status with other demographic, educational, and institutional characteristics. To identify any differential impacts of the pandemic on potentially disparate outcomes, we must first quantify the average impact of the pandemic on EL proficiency, as well as any existing disparities between EL subgroups before the pandemic. As these estimates become available, the impact of the pandemic on the disparities can be assessed by comparing the respective differences in the estimated relationships before and after the pandemic. Therefore, the research questions on differences in average proficiency due to the pandemic and other individual and aggregate factors are as follows:

- **RQ1:** What was the impact of the COVID-19 pandemic on English Learners' average proficiency in the context of (controlling for) individual- and institutional-level factors?

  Individual-level factors:

  - Disparities across *ethnicity (Hispanic)* and *race (for Asian, African/Black, Native American, Pacific Islander or Native Hawaiian, Mixed or Multiple Races, or No Race Reported/Missing);*
  - Impacts of *Time as EL;* differences by *Newcomer, LTEL, SLIFE, Gender, IEP* and *Migrant Status,* and *LIEP Refusal;*
  - Intersections of *Hispanic and Newcomer, Hispanic and LTEL, Hispanic and Female, Hispanic and IEP, Hispanic and Migrant, and Hispanic and LIEP.*

  Institutional-level factors:

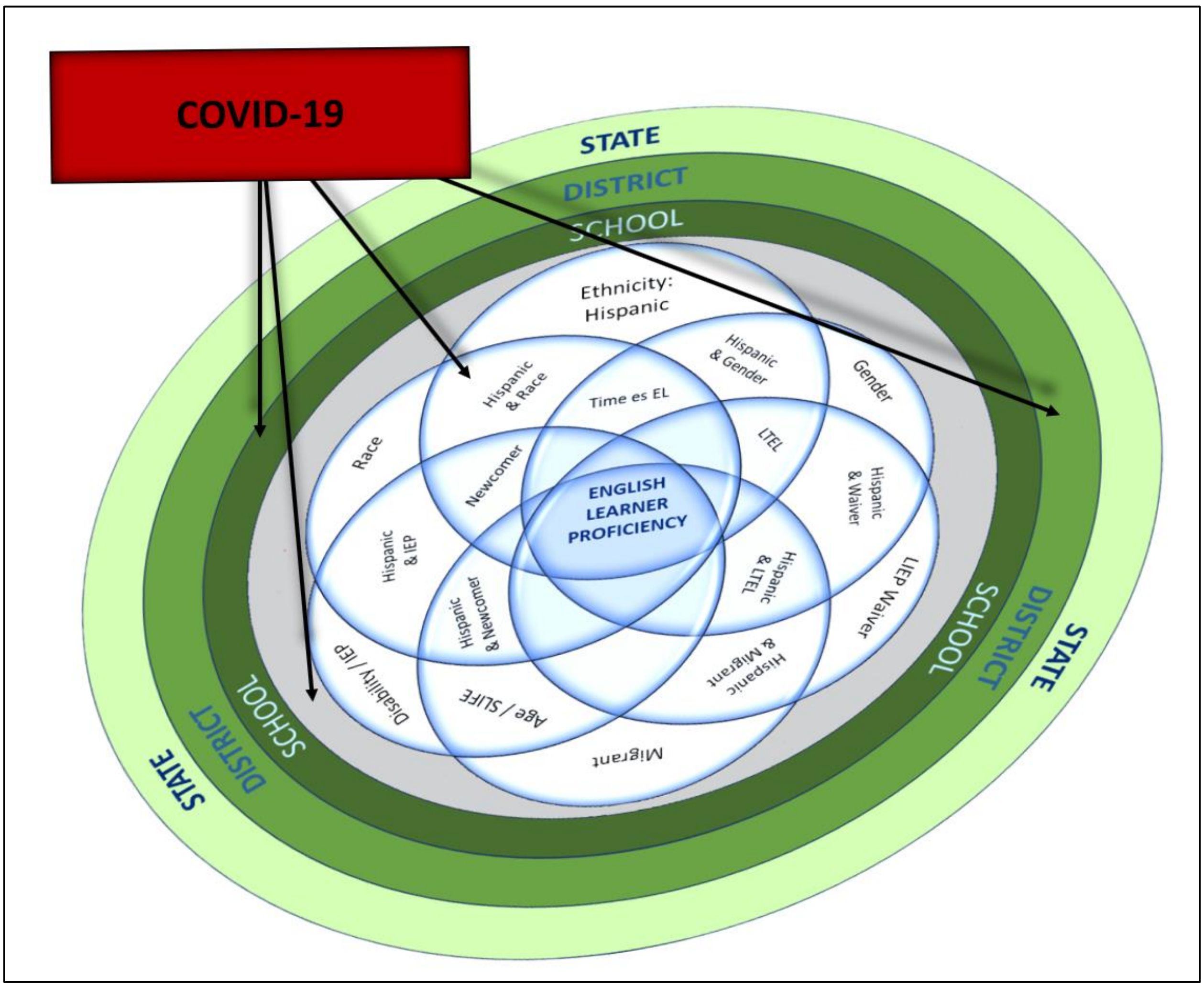
  - Variations in proficiency due to *Schools*, *Districts*, and *States*.

- **RQ2:** How did the COVID-19 pandemic impact each of these disparities and factors?

Conceptual Model

The conceptual model shown in Figure 3.1 is closely aligned with the theoretical framework of *Intersectionality* of English Learners presented in Figure 2.1. It applies an EL subgroup-centered approach in considering the impact of the COVID-19 pandemic on students' proficiency within the context of individual, temporal, and institutional factors available for empirical analysis through the unique dataset described in the next chapter.

Figure 3.1. Conceptual model: Impact of COVID-19 on EL Proficiency in the context of Intersectional identities and institutional-level factors.



Some of the omitted classroom-level variables such as teachers, programs, and peers could be viewed as institutional-level factors that are further nested within schools, districts, and states, grouping sets of students across more levels of structural hierarchies.[15] However, the lack of data on these variables in the WIDA dataset forces me to leave their impact on ELs' proficiency unidentified and unexplored, and therefore (potentially partially) "absorbed" among other included covariates at the individual and institutional levels. Moreover, there are other potentially important, yet not measured and omitted individual-level variables, such as SES, family composition and education that likely impact students' proficiency outcomes across and within various subgroups. Despite these potentially omitted and absorbed factors, an empirical examination of EL proficiency that is based on longitudinal data for the entire population ELs in WIDA states and leverages an *Intersectionality* lens can inform policymakers and educators on important disparities in outcomes. Importantly, findings from this empirical inquiry can highlight differences in ways various student groups are being underserved by an educational system that has been rendered even more ineffective – and in an especially amplified way for specific student subgroups – by the COVID-19 pandemic.

The analysis of these multi-faceted research questions is complicated by the incredible diversity of the English Learner population enrolled in public K-12 schools across the 34 WIDA states included in the analysis. The vast scale of data on millions of students examined across diverse geographic and demographic contexts, and the large number of individual-level covariates/subgroups and their interactions add further

[15] From an assessment design, or psychometric perspective, grade-, or cluster-level effects could also be considered as a level of nesting, further grouping students into test forms of different difficulty. This is left as an area for future research.

complexity to the analysis. Given the different and dynamic ways school-, district-, and state-level policies interact with various components of ELs' education and affect their educational outcomes, addressing the research questions require large amounts of reliable and consistent data across contexts and time, a step-by-step analytic process, and iterative reflection. The next sections present these components.

Data

In this section I describe the source of the data and the variables included in the study. To begin, I provide some geographic and historical context for the population of students taking the ACCESS English language proficiency assessment across the WIDA Consortium and present a short description of how English Learners' test scores and other demographic data are measured, collected, merged, and stored in the ACCESS Longitudinal Dataset. Next, I outline the data inclusion and exclusion criteria, provide detailed information on the dependent variable measuring students' overall proficiency, and the independent variables that are central to the analyses. I provide connections with some of the emerging empirical evidence and outline observed differences and similarities in general trends. Tables and figures examining students' performance across years and subgroups of research interest are given, with a specific focus on average differences in pre- and post-pandemic outcomes measuring student proficiency. Next, I present descriptive evidence of average subgroup disparities in the proficiency outcomes between EL students identified as Hispanic and their non-Hispanic EL peers. The chapter concludes with a presentation of the distribution of students across ethno-racial categories using the interaction of students' ethnicity and race for a more nuanced and accurate identification and estimation of disparities across intersectional EL identities.

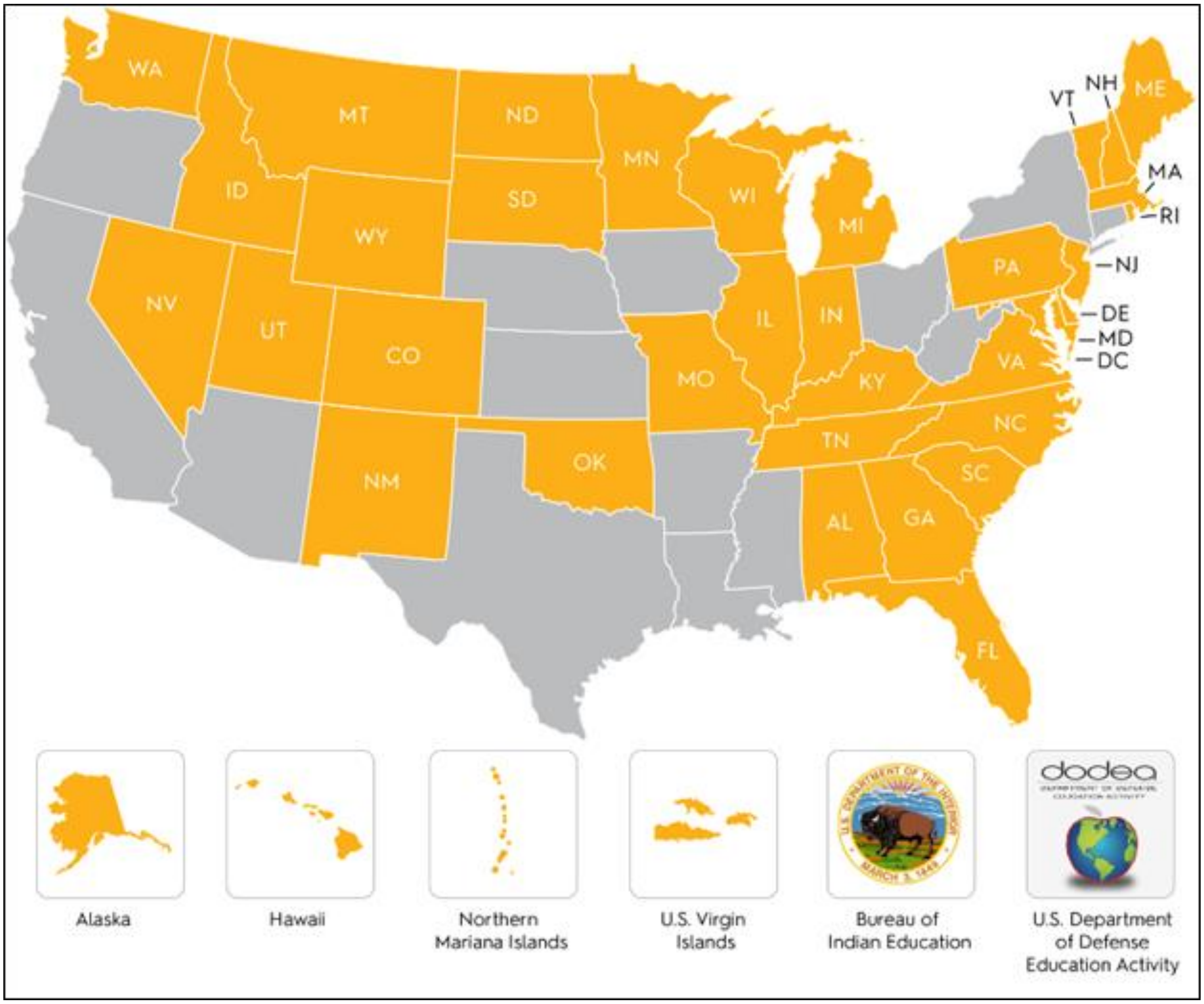
<u>WIDA and ACCESS for ELs</u>

Almost all the empirical research on English Learners opens by reporting that about 10%, or roughly five million EL students are annually enrolled in US schools. This study examines English proficiency outcomes reported by about a quarter of the national population, annually enrolled in schools across WIDA Consortium (WIDA) states in the period spanning 2017–2023.

The WIDA Consortium is currently made up of 41 U.S. states, territories and federal agencies dedicated to the research, design, and implementation of a high-quality, culturally, and linguistically appropriate system of standards and assessments that is intended to support English Learners in K-12 contexts (Figure 3.2). WIDA was established in 2003 after the authorization of NCLB (2001), when an Enhanced Assessment Grant (EAG) was awarded to the Wisconsin Department of Public Instruction, WIDA's first home.[16]

[16] The name WIDA originally stood for the four states on the grant proposal: Wisconsin, Illinois, Delaware and Arkansas. Today, the name WIDA has come to represent the entire WIDA Community of states, territories, and federal agencies.

Figure 3.2. Map: WIDA Consortium Member States, 2023.[17]



Funded by the grant, WIDA developed the 2004 WIDA English Language Proficiency Standards, which served as the basis for the ACCESS for ELLs assessment of English Language Proficiency. These standards were aligned to the academic content standards of the members of the WIDA Consortium and adopted by Teachers of English Speakers of Other Languages (TESOL) (Fox & Fainbairn, 2011). Based on these

[17] Map adapted from: https://wida.wisc.edu/memberships/consortium. Washington became a member of the consortium in 2021 and is not included in the study. Northern Mariana Islands, US Virgin Islands, Bureau of Indian Education and US DoDEA were excluded from the analytic sample due to substantial rates of (non-randomly) missing demographic data, or other data on ethno-racial demographic information that was focal to this analysis.

standards, ACCESS for ELLs® was launched in 2005 under the direction of the Center for Applied Linguistics (CAL), the principal developer of the assessment. In 2006, WIDA moved to its current home at the University of Wisconsin–Madison, within the Wisconsin Center for Education Research. At UW-Madison, WIDA expanded and improved its comprehensive system of assessments, the core one being its suite of large-scale English language proficiency tests for K–12 students: ACCESS for ELLs. It is a central component of WIDA's comprehensive, standards-driven system that supports the teaching and learning of ELs. ACCESS is a standards-referenced test, which means that student performance is compared to English language development standards WIDA has defined. Important for the design of the study, performance is not capped; any student can achieve any score, in any given year. Students' performance is not ranked against each other, or against the expected performance of monolingual English speakers.

The ACCESS assessment is not mandated by federal legislation; however, it meets key federal requirements related to the education of English learners.[18] ACCESS is intended to assess reliably and validly the English language development of English learners in Grades K–12 (WIDA Consortium, 2012). One of the purposes of ACCESS is to monitor student progress in English language proficiency on a yearly basis, and to serve as one of the criteria that educators in WIDA states consider as they determine whether English Learners have attained an English language proficiency level that will

[18] More specifically, under ESSA (2015), states are required to establish English Language Proficiency (ELP) standards and assessments to measure the progress of English Learners in acquiring English language proficiency. ESSA, Section 1111(b)(2)(G). The ACCESS for ELLs Online assessment, developed by WIDA in 2016, has been used in WIDA Consortium states to fulfill this requirement.

allow them to meaningfully participate in English language classroom instruction.[19] The design, configuration, review, and administration of the ACCESS annual high-stakes English language proficiency assessment is an immense effort and is coordinated between WIDA, housed at the UW-Madison; an assessment vendor (Data Recognition Corporation) headquartered in Minneapolis, MN; the Center for Applied Linguistics (CAL), located in Washington, DC, as well as thousands of school districts in a consortium of over 40 state educational agencies.[20] A rigorous and detailed 750-page technical report is provided annually, describing the psychometric methods, analytic processes, and operational steps undertaken to ensure a high-quality, reliable, and consistent assessment instrument (WIDA Annual Technical Report for ACCESS for ELLs Online English Language Proficiency Test Series 503, 2021-2022 Administration).[21]

Further, a Bias and Sensitivity Review Panel ensures that test items and tasks are free of material that might favor any subgroup of students over another on the basis on gender, race/ethnicity, home language, religion, culture, region, or socioeconomic status, and/or be upsetting to students. Additionally, CAL uses differential item functioning (DIF) analysis to investigate whether factors extraneous to English language proficiency (i.e., the construct being measured on the test) may have influenced some students' performances on items. DIF attempts to find and filter out test items that may be

---

[19] Overall composite proficiency levels (computed based on the overall composite scale scores) at which ELs are considered for reclassification vary across states; albeit the degree of the variability in reclassification criteria has decreased since the adoption of the ACCESS for ELLs Online assessment across WIDA states in 2016.

[20] Further, the Consortium provides online test administrator training courses, sample items, tasks, and rubrics in order to facilitate classroom activities tied to the standards and representative of the kinds of language production expected on the ACCESS for ELLs.

[21] WIDA Annual Technical Report for ACCESS for ELLs Online English Language Proficiency Test Series 503, 2021–2022 Administration, (2022); Annual Report No 18A. Prepared by the Center for Applied Linguistics; Retrieved from: https://wida.wisc.edu/sites/default/files/resource/ACCESS-Online-ATR-2021-22-redacted.pdf

functioning differently than intended for specific student subgroups. Importantly, and pertinent for the methods applied in this report, DIF analyses are implemented for the ethnicity and gender variables, targeted at reducing differential item functioning for *Hispanic* vs *non-Hispanic* and *Female* vs *Male* students.

In sum, the ACCESS Online assessment is a reliable, multi-stage, semi stage-adaptive test, based on a modified linear Rasch (1960) model.[22] The above-discussed reviews, administrative processes, and psychometric analyses are aimed at ensuring a high-quality, reliable, consistent, valid, and equitable language proficiency assessment. As such, WIDA's ACCESS for ELLs Assessment has been used nationally across over 40 states and US territories (as well as internationally) over the last two decades and has been hailed as the standard in English language assessment by many renowned researchers in the field (Fox & Fairburn, 2011; Kenyon et al, 2007). The data underlying the analysis is based on assessment scores and demographic data on all students identified as English Learners in WIDA states. The next section provides a brief description of the ACCESS longitudinal dataset containing the variables of research interest.

[22] In a fully computer-adapted test a subsequent item is easier or harder depending on answer accuracy. A drawback to this type of approach is that it requires extended testing sessions to have sufficient reliability (MacGregor et al., 2022). WIDA's multi-stage adaptive (MST) framework allows for sets of questions to be analyzed together, and based on student performance, offer up the next module (with subsequent sets of questions). Technical statistics on the reliability, consistency, and validity of the test can be found in the technical report provided in the previous footnote.

ACCESS Longitudinal Dataset [23]

The analytic sample consists of test scores from the entire population of English Learners who have taken the ACCESS for ELLs Online (ACCESS) assessment across WIDA states and includes some demographic and assessment-related data on English Learner students taking the assessment. Inspired by the *Intersectionality* lens and aiming for generalizability of findings across the WIDA Consortium, I cast the widest possible net on the scale and scope of the data across space and time, including as many school years (2017–2023), geographic locations (34 WIDA states), and demographic variables as the quality of the reported data allows without compromising the research design. Due to differences in how some of the data is reported and collected, this required finding a balance between the scale and the scope of the data. For example, the analytic sample does not include ELs in kindergarten, where the ACCESS assessment is administered as a paper test. Even though including the paper test would further bolster the sample size and likely increase the already extremely high estimated precision parameter estimates, this would come at the cost of losing several variables of interest, such as students' age and the duration identified as an English Learner, which are not recorded as accurately in the paper version of the assessment.[24]

Many other student-level demographic variables are recorded in the ACCESS longitudinal dataset (such as students' native language, type of LIEP program, 504 status, etc.). However, due to inconsistent reporting across states and other factors affecting

[23] The analysis contains no identifiable state-, district-, school- or individual-level information. This research has been approved by the University of Wisconsin-Madison Education and Social/Behavioral Science Institution Review Board (IRB # 2013-0558) in accordance with federal regulations, state laws, and local and University policies.

[24] Sensitivity analysis revealed that including the paper test results in the analysis does not substantially alter the findings regarding the average impact of COVID-19, as well as on subgroup disparities.

data quality, these variables could not be included in the study. Future research will examine the feasibility of analyses focused on these variables, potentially based on data from more localized settings where data may be of higher quality. Further, while WIDA started administering ACCESS as an online assessment in 2016, data from this school year is excluded from the study, as a) this was the first year an online assessment was administered across the Consortium, b) several states joined WIDA in 2017, c) a new standard-setting was conducted in 2016, reflecting the higher rigor of Common Core standards (Sato & Thompson, 2020) on 2016-2017, and subsequent ACCESS scores.

The 34 WIDA states included in this study were part of the Consortium throughout the period examined in this study spanning the school years 2016–17 (SY 2017) to 2022–23 (SY 2023). Altogether, the analytic sample includes just under three and a half million EL students' records of measured English proficiency across a period of multiple years, spanning both pre- and post- COVID-19 academic school years. Students' records are connected longitudinally, providing just under ten million student-by-year observations of English proficiency, as measured by *Overall Composite Scale Scores*, for all students identified as English Language Learners enrolled in grades 1–12 across the WIDA consortium. Therefore, for each school year considered (2017–2023), all students included in the study are "active" ELs by design because they took the ACCESS assessment at least once.[25] In addition to the test score data on English proficiency, individual-level demographic variables such as students' age, gender, race, ethnicity, as IEP status (Individualized Education Plans), migrant status, and waiver status (students

[25] In other words, the analytic sample does not include current data on "former" ELs for a specific year, as they did not take the ACCESS assessment. Also unavailable is data on students who move out of the state or country.

whose parents refuse English language support services) are also available and included in the analysis as independent variables. The data on these variables has two sources: a) the WIDA Assessment Management System, enabling ACCESS test administrators to input data on students' reported demographic information during the annual assessment of their English proficiency, and b) state education departments across the WIDA consortium that receive these data annually, review and correct it if needed, and share it back with the WIDA Consortium based on a Memorandum of Understanding, for additional data validation, psychometric, assessment research, and test development purposes. Each year upon the completion of the ACCESS assessment administration, WIDA checks, processes, and merges these records with existing data from the previous ACCESS administrations according to a matching algorithm based on students' first and last name (removed from research datasets for confidentiality and privacy purposes), birthdate, and other demographic data. Descriptive statistics for these variables are given at the end of the Data section in Tables 3.9 and 3.10.

English Learners' Average Proficiency: Overall Composite Scale Scores

English Learners' proficiency, as measured by their *Overall Composite Scale Scores (CSS*) is the focal variable of interest in this study, shown centered in the middle of the theoretical framework (Figure 2.1) and impacted by COVID-19 in the conceptual model (Figure 3.1). *CSS* is computed as a weighted average of students' scores in four individual language assessments in the domains of reading (35%), speaking (15%), listening (15%) and writing (35%).[26]

The analytic sample, therefore, includes only those students who have completed all four domains of the ACCESS online assessment. Each of the 3,391,969 unique EL students in the dataset reported at least one measurement of *CSS* in the dataset, recording their overall English proficiency at a certain point in time (limited to academic years 2017-2023), as tested in a specific WIDA school, district, and state where their annual ACCESS online assessment took place.[27] Table 3.1 presents the distribution of WIDA's English Learner student population across grades 1-12 and tested in years 2017-2023.

Table 3.1. WIDA's EL population across grades (1-12) and years (2017-2023).

Grade/Year	Pre- COVID-19				Post- COVID-19			Total (Grades)
	2017	2018	2019	2020	*2021**	2022	2023	
1	161,884	173,593	181,012	174,515	*141,732*	181,047	183,803	1,197,586
2	165,545	178,086	185,830	180,923	*144,072*	183,002	181,955	1,219,413
3	173,951	184,300	185,973	178,447	*143,678*	182,746	177,660	1,226,755
4	115,999	175,786	178,699	171,894	*137,797*	179,372	171,210	1,130,757
5	84,876	114,285	142,882	137,687	*107,205*	154,494	140,440	881,869
6	72,837	85,399	102,140	112,808	*84,991*	124,670	121,746	704,591
7	74,444	81,202	91,017	101,836	*85,456*	117,750	122,276	673,981
8	73,644	79,429	84,117	88,073	*75,329*	119,363	114,907	634,862
9	92,048	90,105	96,070	99,133	*64,788*	126,354	131,558	700,056
10	61,764	82,934	79,431	78,015	*57,561*	86,507	106,638	552,850
11	43,395	59,322	70,319	64,028	*46,166*	72,648	75,763	431,641
12	29,257	40,471	50,506	55,073	*34,060*	58,690	61,474	329,531
Total (Years)	1,149,644	1,344,912	1,447,996	1,442,432	*1,122,835*	1,586,643	1,589,430	9,683,892

[26] While the ACCESS assessment is untimed, WIDA suggests the following durations for test sessions: reading (60 minutes), listening (65 minutes), speaking (50 minutes), and writing (90 minutes). The full assessment, therefore, takes an EL student an average of 265 minutes to complete, but can be spread over several days (WIDA, 2021a).
[27] Future research will examine the impact of COVID-19 on students' outcomes in the individual language domains.

Taken annually by all students identified as ELs in WIDA states, the ACCESS assessment was administered “during COVID-19”, i.e. in the academic school year of 2021. However, not only did COVID-19 force closure of schools and transition to remote or hybrid instruction at differing times, to varying degrees, and for differing durations depending on the geographic locale and other socio-economic, political, and biological factors, but the states’, districts’, and schools’ responses to the myriad of challenges brought about by the pandemic were also varied and context-dependent. Such responses included states’ requesting of waivers from the annual ACCESS assessment of students in entirety, or intentionally testing only potentially higher-performing subgroups, or shifting and/or extending the states’ and districts’ otherwise regular annual testing cycles. Altogether, in the 2021 academic school year an estimated third of the EL student population was not tested due to various pandemic-induced reasons, while most, if not all ELs that were tested in 2021 took the ACCESS assessment under unprecedented and irregular circumstances. Despite these factors potentially introducing higher uncertainty and larger measurement errors to the analysis, including students’ test scores from the 2021 school year in the analytic sample provides longitudinal continuity to many of the student records across the seven-year timespan (Table 3.2), while excluding these records does not significantly alter the findings.

Table 3.1 shows that WIDA’s English learner population is disproportionately spread out across grades. Most of EL students are enrolled in earlier, elementary-level grades, as every year new students enroll in US schools and are identified as English learners, while (fewer) others who received high scores on their previous ACCESS test are reclassified (sometimes subject to additional state-defined criteria), exiting both EL status and the ACCESS longitudinal dataset before reaching middle or high school. Table 3.1 also shows that but for the substantial dip in the academic year of 2020–21 when the pandemic began, the number of tested ELs has gradually increased from a total of about 1.1 million in 2017 to about 1.5 million in 2023. This observed increase in the overall English Learner population has to do with national demographic patterns and immigration trends and is likely bolstered by the growing numbers of English Learners who continue to be identified as ELs for extended, and increasingly longer periods of time. [28]

Vertical Scaling of CSS and Grade Fixed effects

Overall *Composite Scale Scores* (CSS) range from 100 to 600 and are vertically scaled, thereby enabling comparisons of students' proficiency across different grades.[29] These vertical scale scores are used to compare "equivalent knowledge across grades", as well as to monitor an individual student's yearly growth (WIDA, 2022, p. 5). In other words, a *CSS* of 300 is calibrated for, and intended to reflect a "similar" level of language proficiency, regardless of the grade the student is attending. Furthermore, the ACCESS Online assessment is grade-level cluster-based (1-2, 3-5, 6-8, and 9-12), so ELs within specific grade clusters take (computer-adaptive) test forms of similar difficulty.

These factors ameliorate potential issues with respect to comparability of scale scores representing English proficiency of EL students in different grades from a statistical validity standpoint. However, due to the cumulative nature of the process of language acquisition, and the relatively large differences in average composite scale

[28] As stipulated by federal non-regulatory guidance, after five years in a language support program, these English Learners should be accounted for and reported by schools and districts as *long-term ELs*.

[29] Vertical scaling of scores is accomplished by an equipercentile linking process, with grade 6 scores centered at the middle of the distribution (ref).

scores across grades reported in Table 2, the descriptive data and graphical evidence is aggregated by grade, while the regression analysis described in the next section is performed using grade fixed-effects. This approach makes sure that students' outcomes are effectively being compared within grades, or more precisely, taking into account average grade-level differences that may be inherent to the English language development and measurement processes, independent of the vertical scaling.

CSS: Rounding and Formatting

In the WIDA dataset, CSS are constructed though a weighted average of students' performance in four individual language domains of reading, speaking, listening, and writing. Overall CSS points are rounded to the closest integer and reported as a single point within a confidence band, termed the "conditional standard error of measurement" (CSEM). This designates a single point as the smallest reported unit of difference between scores at the individual, student-level. Scale scores allow the difficulty of items to be measured using a common test construct, resulting in correlated scale scores across tests and across kindergarten through 12th grades (Gottlieb & Kenyon, 2006; Gottlieb et al., 2007).

Moreover, in any given year for many English learner students taking ACCESS a difference of a single CSS point in measured proficiency can mean the difference between the attainment of (state-defined) reclassification-level proficiency and subsequent exit from EL status, or conversely, another year in language support services, with another ACCESS test upcoming in the next academic year. Due to these reasons, leveraging the tremendously large sample sizes resulting in very precise estimates, in the reporting of descriptive statistics and parameter estimates of regression models, I also report CSS points rounded to the nearest integer. This approach achieves better legibility in the comparisons of results from multiple models and dozens of covariates without loss of nuance or generality and draws attention to more meaningful changes and trends, or lack in thereof. Decimal points are shown only for the focal parameters of research interest, and when highlighting notable differences across various model specifications or subgroup outcomes. [30] For similar purposes, in presentation of graphical evidence and descriptive trends informing the research questions on the impact of COVID-19 on EL proficiency subgroup disparities, I use conditional formatting. Regardless of the outcome being examined, in this and following tables red, yellow, and green shading is used to indicate relatively “low”, “medium” and “high” numbers, respectively. This type of representation is superior to line or bar charts, which can become very busy and overwhelming when presenting data for 12 separate grades, multiple subgroups of students, and several academic school years.

CSS: Descriptive Statistics

Table 3.2 presents descriptive statistics on ELs’ average scale scores using conditional formatting (based on grade).

[30] I also show decimal points when the estimate would otherwise be rounded and show ‘0’. Precise estimates are provided in Appendix A.

Table 3.2. ELs' average proficiency across grades (2017-2023 average).

Grade	Mean	Std. dev.	Min	Max	Freq.
1	274	33	120	409	1,197,586
2	300	33	129	414	1,219,413
3	318	36	146	423	1,226,755
4	347	35	145	465	1,130,757
5	352	38	155	489	881,869
6	337	33	204	476	704,591
7	343	36	189	470	673,981
8	349	39	196	494	634,862
9	354	38	224	493	700,056
10	361	36	211	490	552,850
11	367	35	211	493	431,641
12	368	34	227	495	329,531
Average	331	46	-	-	9,683,892

The first column in Table 3.2 shows that the proficiency of English learner students increases from an average CSS of 274 in grade 1 to 368 in grade 12, as active ELs (i.e., those who do not get reclassified or drop out) progress through the grades and acquire higher levels of academic English. The dip in average proficiency in middle grade schools, reverting the monotonicity in the increase, is likely due to the large number of higher-proficiency students exiting the EL status (and the analytic sample) before middle school, as well as the jump in academic expectations in middle school as compared to elementary school, which is also reflected in ACCESS scaling. Despite the aggregation across time, standard deviations are relatively stable across grades, ranging from 33 (in grade 1) to 39 overall composite scale score points (in grade 8).

Table 3.3 presents EL's average proficiency across time, for the school years 2017-2023, and grades 1-12.

Table 3.3. EL's average overall composite scale scores by grades school years.[31]

Grade / Year	Pre- COVID-19				Pre-Covid Average	Post- COVID-19			Post-Covid Average	Impact of COVID-19
	2017	2018	2019	2020		*2021**	2022	2023		
1	283	282	279	277	280	*270*	264	265	266	-14
2	302	306	304	304	304	*298*	293	295	295	-9
3	320	323	322	322	322	*316*	312	311	313	-9
4	342	352	351	351	349	*344*	344	342	343	-6
5	347	354	357	356	354	*349*	351	347	349	-5
6	336	339	341	340	339	*335*	337	333	335	-4
7	344	345	345	344	345	*343*	341	340	341	-3
8	351	351	350	349	350	*349*	348	344	347	-4
9	355	359	358	352	356	*359*	351	351	352	-3
10	359	366	365	362	364	*361*	360	356	359	-5
11	364	370	371	369	369	*368*	364	362	364	-5
12	368	371	371	370	370	*370*	367	363	366	-4
Average	329	334	335	334	333	*329*	329	327	328	-5

The conditional formatting capturing across-time differences in average proficiency for each of the grades depicts consistent patterns of a large and sustained impact of the pandemic on average proficiency after the 2020 academic school year, for each of the grades 1-12. Importantly, Table 3 provides descriptive evidence that there is little variation in average EL proficiency across years when pre- and post-pandemic years are considered separately. Annual differences in average proficiency have been rather consistent within grades, ranging from 1 (in grade 7) to 9 (in grades 4 and 5) *CSS* and exhibiting an increasing trend before the pandemic and a decreasing trend after the pandemic. This within-grade consistency of scores across time enables descriptive comparisons of pre- and post-COVID-19 averages, which confirm a substantial impact of the pandemic for each of the grades, presented in the last column of the table. Notably, due to the impacted sample in 2021 missing about 30% of observations (not at random)

[31] Conditional formatting in this table captures differences in scale scores temporally (across years).

due to the ongoing pandemic, the post-pandemic averages are likely overestimates, while the COVID-19 impact estimates (pre-post differences in average scales scores) presented in the last column are underestimates. The aggregate (averaged across grades), impact of the pandemic on ELs' proficiency is estimated at about -5 *CSS*, with the larger learning losses reported in earlier grades (a decline of 14, 9 and 9 *CSS* in grades 1, 2 and 3), enrolling the largest proportion of the EL population. Average declines in scores were smaller in grades 4-12, ranging from 3 to 6 *CSS*.

Finally, and perhaps most worryingly, the estimates highlighted in dark red shading in the last column of Table 3.3 call attention to the fact that English Learners' average proficiency is still declining, and for the most recent assessment of students in 2023 was at its recorded lowest since 2017.

<u>Disparities in Hispanic vs non-Hispanic EL Proficiency</u>

Next, as preliminary descriptive evidence of hypothesized subgroup disparities and how they were impacted by the pandemic, Table 3.4 provides a first view at by-grade and by-year differences in average proficiency by *Ethnicity* identification. More specifically, it juxtaposes *Hispanic* and *non-Hispanic* English Learner students' outcomes.

Table 3.4. Subgroup disparities between Hispanic and non-Hispanic ELs by grade and years.

Grade / Year	Disparity Pre-COVID-19				Pre-COVID-19 Hispanic Disparity	Disparity Post-COVID-19			Post-COVID-19 Hispanic Disparity	Impact of COVID-19 on Hispanic Disparity
	2017	2018	2019	2020		*2021**	2022	2023		
1	-9	-10	-10	-13	-11	-16	-17	-16	-15	-4
2	-7	-8	-9	-12	-9	-14	-15	-14	-13	-4
3	-5	-7	-7	-10	-7	-11	-13	-13	-11	-4
4	0	-4	-4	-6	-4	-9	-9	-8	-8	-4
5	0	-2	-1	-5	-2	-7	-7	-6	-5	-3
6	-2	-3	-2	-5	-3	-5	-5	-5	-5	-1
7	-4	-5	-5	-7	-5	-7	-7	-7	-6	-1
8	-4	-5	-5	-9	-6	-8	-8	-7	-7	-1
9	-6	-7	-6	-13	-8	-7	-9	-9	-8	0
10	-8	-8	-7	-9	-8	-9	-10	-10	-9	-1
11	-9	-8	-5	-6	-7	-7	-10	-10	-8	-1
12	-6	-8	-5	-4	-6	-3	-8	-10	-7	-1
Average	-5	-6	-5	-7	-6	-7	-8	-8	-7	-2

The last column of Table 3.4 presents the pre- vs post- COVID-19 disparities in the average proficiency of *Hispanic* versus *non-Hispanic* identified English Learners for each of the grades 1-12 and aggregated across grades in the last row. Table 3.4 shows that while disparities between *Hispanic* and *non-Hispanic* students' proficiency existed in each of the grades 1-12 prior to the pandemic, these disparities increased after the onset of the pandemic with the exception of grade 9. Interestingly, the data indicates that average disparities by ethnicity increased even in 2020, warranting a more rigorous examination of these descriptive aggregate trends. Overall, the disparity between Hispanic and non-Hispanic students (averaged across grades) has slightly increased, with the larger disparities reported in elementary school grades (1-5).

These comparisons of average disparities across ethnic subgroups are estimated without regard for the students' race. In other words, the outcomes of ELs who are

Hispanic and *White* are grouped with those who are *Hispanic* and *Asian*, and *Hispanic* and *Native American,* etc. However, while this oversimplification requests further unpacking, it cannot be remedied though dozens of tables disaggregated both by ethnicity and race. Instead, I address the intersection between ethnicity and race through the use of regression models (described in the section on the analytic strategy), which are better suited to exploring relationships among large numbers of variables.

In sum, the descriptive evidence presented in Tables 3.3 and 3.4 provides preliminary evidence of a large and differential impact of the pandemic on EL's average proficiency. However, it is important to recall that while the average estimates are based on millions of observations and are thus reliable in describing aggregate trends, they are also unconditional estimates (except for the grouping by grades, school years, and ELs' identification as *Hispanic*). In other words, these descriptive by-grade and by-year averages, while informative, could mask potentially important impacts of individual- and institutional-level factors which are likely relevant for the very diverse population of English Learner students.

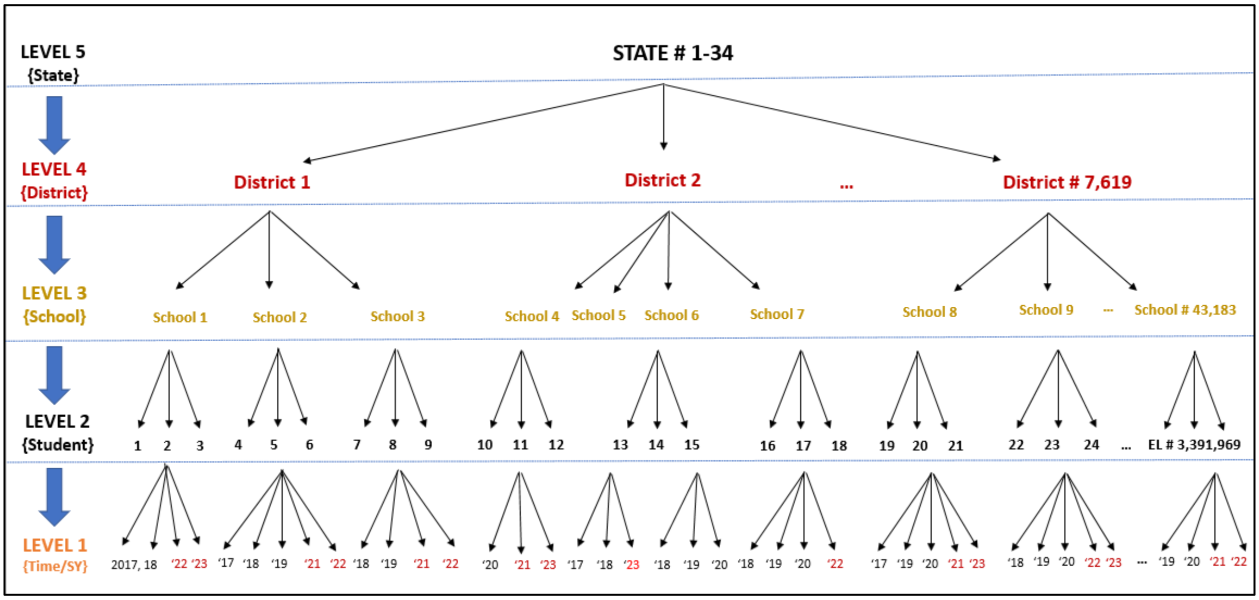
State, District, School, and Student Identifiers

In addition to test scores measuring students' English language proficiency in a given academic year, state, district, school, and student identifiers (district, school, and student numbers) are available for each individual test record, providing complete information on WIDA's EL ≈ 3.3 million students' enrollment and nesting within 43,183 schools, 7,619 districts and 34 states where (and when) the ACCESS annual English language assessment was administered. State, district, school, and student numbers are de-identified, but unique and consistent across time, thereby enabling both longitudinal and hierarchal examinations of students' test scores. Having complete nesting information on each student with respect to the state, district, and school in which they were enrolled when they took the ACCESS test addresses some of the issues in examinations of average proficiency that are due to the tremendous heterogeneity in English learner students' clustering across and within these different levels. For example, there are many schools and districts reporting just a handful of tested English learner students. There are states that have only a few districts, while there are others that have hundreds. Further, there are many districts that have only one school (with enrolled and tested EL students), and many more that have dozens. The analytic strategy leverages this nesting of students within schools, districts, states, as well as the repeated observations for within students across time, to decompose and quantify variations in average English proficiency that are related to each of these factors through random-effects specified at each of the four levels.[32]

Figure 3.4 provides a visual description of the nested structure of the data, with only one state shown for legibility. For similar purposes, the figure also does not depict the cross-nested nature of the data when considered temporally: many EL students change schools and districts throughout their academic journey, both as a part of regular transition from elementary to middle to high schools, but also due to moving. The section on analytic strategy provides more details on how the mixed-effects regression models handle this nuance.

[32] More granular data with classroom and teacher identifiers could potentially provide a more complete nesting and enhance the ability to further decompose variations in student performance, for instance by classroom teacher. However, absence of such data, as well as issues with small(er) samples (at the smallest cluster-levels) and problems with attribution (of students' scores to specific teachers) arise, as discussed in Sahakyan and Cook (2009).

Figure 3.3. Five-level structure of the data: States → Districts → Schools → Students → Time



Time / School Years

The timespan of this study includes the academic years 2017 through 2023, providing seven school years of annual assessment data on the entire population of English learners tested online throughout this time-period (Table 3.1). Supported by the underlying large sample sizes, the included timespan appears relatively long and potentially sufficient with respect to researchers' ability to accurately quantify and model long-term language acquisition process for EL students. [33] However, several factors complicate approaches that might focus more on estimating long-term growth trajectories of students over time. One of these factors is the high mobility of the EL student population being examined, as compared to their non-EL peers. For example, every school year many students new to the US are identified as "newcomer" English learners and take the

---

[33] Prior research suggests that while there are large contextual differences by students, classroom and school factors (Slama et al, 2014), on average it takes 5-7 years to reach academic English proficiency (Linquanti & Cook, 2015; NASEM, 2017).

ACCESS test for the first time, while others, already identified and enrolled in (English) language instructional educational programs (LIEPs) reach sufficiently high levels of English proficiency and exit both EL status and the analytic dataset. Other students drop out of programs and schools before reaching state-defined reclassification-level English proficiency, or move out of the state or country.[34] Moreover, some of these students, (even a few among those who had previously been reclassified), enroll in the same, or another state and/or school, and take the ACCESS assessment again, and are recaptured in the ACCESS dataset.[35] This high mobility of EL students is recorded in the ACCESS longitudinal dataset though intermittently missing observations in the fields measuring students' annual test scores on overall proficiency, i.e., overall *Composite Scale Scores (CSS)*. Therefore, a missing observation of *CSS* for an EL student in a particular year could mean that the student has either been reclassified (based on their high score in their previous year's ACCESS administration), or dropped out of school, or moved away (out of state or country), or otherwise was not able to take (all four domains of) the ACCESS assessment.

Reflecting this high mobility, Table 3.5 presents, in order of decreasing frequency, the top 50 longitudinal patterns of assessment data that have been recorded for the ≈ 3.4 million English learner students throughout the 2017–2023 timespan. Missing data on

---

[34] WIDA is unable to track students who move across states. A matching algorithm is used to track students who move within states, across districts and schools. EL students who move across WIDA states and take the ACCESS test again are assigned a new unique identifier by the state and are thus counted newcomer students.

[35] Federal legislation requires that school districts must monitor the academic progress of former EL students for at least two years to ensure that students have not been prematurely exited; any academic deficits they incurred resulting from the EL program have been remedied; and they are meaningfully participating in the district's educational programs comparable to their peers who were never EL students (never-EL peers). https://www2.ed.gov/about/offices/list/ocr/docs/dcl-factsheet-el-students-201501.pdf

students' overall composite proficiency (CSS) are denoted with a dot, while '1' indicates the presence of a valid *CSS*. The presence of pre-COVID-19 observations is identified by the number 1 in white, for the first four academic years, while post-COVID scores are marked in red, for the three years after 2020. For example, according to the figures reported in Table 3.5, the highest frequency pattern is observed for about 12% of students reporting assessment data only in 2023 (i.e., *Newcomer* ELs), while an additional 9% report scores in both 2022 and 2023 school years.

Table 3.5. Longitudinal patterns in ACCESS test-taking: School Years 2017-2023.

↓ #	CSS Pattern: '17 '18 '19 '20 '21 '22 '23	%	Cumulative %	Freqeuncy	↓ #	CSS Pattern: '17 '18 '19 '20 '21 '22 '23	%	Cumulative %	Freqeuncy
1	1	11.9	11.9	402,122	27	...11..	1.1	88.9	35,979
2	11	9.4	21.2	318,472	28	.11111.	1.0	89.9	34,971
3	1......	6.1	27.3	206,334	29	.1111..	1.0	90.9	33,446
4	11.....	5.8	33.1	196,271	30	..11.11	1.0	91.9	32,561
5	111....	5.1	38.2	172,821	31	..111..	0.9	92.8	31,493
6	111	4.5	42.7	152,403	32	.111.11	0.8	93.6	27,881
7	1111111	4.5	47.2	151,831	33	1111.1.	0.8	94.4	25,562
8	1111...	4.2	51.4	143,621	34	..11.1.	0.5	94.8	15,866
9	...1111	4.1	55.5	138,991	35	...1.1.	0.5	95.3	15,142
10	.....1.	2.9	58.4	98,330	36	.111.1.	0.4	95.7	13,208
11	..11111	2.9	61.3	98,202	37	111.111	0.3	95.9	8,745
12	.1.....	2.6	63.9	87,212	38	1.1....	0.2	96.1	7,225
13	..1....	2.4	66.2	80,248	39	..1.111	0.2	96.4	7,138
14	...1...	2.3	68.5	78,081	40	111..11	0.2	96.6	6,861
15	.111111	2.2	70.8	76,012	41	11.1...	0.2	96.7	5,430
16	.11....	2.0	72.8	67,065	42	111..1.	0.2	96.9	5,232
17	111111	1.8	74.6	61,476	43	1.11...	0.2	97.0	4,976
18	1111.11	1.8	76.4	60,185	44	11111.1	0.1	97.2	4,780
19	..11...	1.7	78.1	59,067	45	....1.1	0.1	97.3	4,460
20	.111...	1.7	79.8	57,718	46	..1..11	0.1	97.4	4,441
21	11111..	1.5	81.3	50,199	47	.11.111	0.1	97.6	4,432
22	..1111.	1.5	82.7	49,322	48	1.11111	0.1	97.7	4,168
23	...1.11	1.4	84.1	47,907	49	.11..11	0.1	97.8	4,074
24	...111.	1.3	85.5	45,490	50	111.1..	0.1	97.9	4,039
25	....11.	1.2	86.7	40,783	51-127	All Other	2.1	100.0	70,750
26	....1..	1.2	87.8	38,946	TOTAL	XXXXXXX	100	100	3,391,969

The longitudinal patterns of test-taking reported in Table 3.5 also make it clear that restricting the analytic sample to only those ELs who have non-missing language assessment data across multiple adjacent years would drastically reduce the sample size. For example, if the analysis were limited to only those English learner students who have non-missing observations (of *CSS*) throughout the entire timespan of the study (pattern #5), this would imply a loss of about 95% of the sample. In addition to issues with much smaller sample sizes, limiting the sample to students with assessment data across multiple adjacent years would also inadvertently shift the focus of the study to "long-term" ELs.

The high mobility inherent within the EL population, further exacerbated by the pandemic's impact on ELs' assessment in 2021 resulting in a substantially lower number of tested students across pandemic-adjacent years, lends additional support to the selected analytic strategy of examining ELs' English proficiency as measured in a given year – while additionally controlling for potential individual, temporal, and other effects, as explained below – rather than long-term language acquisition over time. Examination of EL student growth, i.e. individual students' acquisition of English language proficiency across time, is left as an area for future research. [36]

Time as English Learner

Frequently cited research conducted by Hakuta, Butler, and Witt (2000) examined cross-sectional data and found that most students take between 2 and 5 years to develop oral English proficiency, and 4 to 7 years to achieve English Language Arts (ELA)

[36] Instead, the longitudinal structure of the data and the repeated observations across time for are modeled though student-level random effects and autoregressive (AR1) residuals. (See the Analytical strategy section. I also test and present a longitudinal specification – Model 3 – for sensitivity analyses and robustness checks).

proficiency. More recent analyses using survival analysis methods and longitudinal data, such as those by Conger (2009), Thompson (2015a), and Umansky and Reardon (2014), have been employed to predict how much time students typically need to meet English proficiency benchmarks. While using similar analytical techniques, these studies have shown differing durations, influenced by the varied reclassification criteria set by different school districts. Specifically, Conger (2009) reports that in New York City the median duration to attain the required ELP was around 3 years. Research in two large urban California districts (Thompson, 2015a; Umansky & Reardon, 2014), which factored in extra benchmarks for reclassification, revealed that the average time to exiting EL status extended to about 6 to 6.5 years, while over a quarter of students were not reclassified even after 9 years. Recent research examining data from WIDA corroborates these variations in time to reclassification and their relationship to both individual-level factors (Sahakyan & Poole, 2023; Sahakyan & Poole, 2021) and varying reclassification criteria across states and districts (Sahakyan & Ryan, 2018).

A major advantage of the unique longitudinal dataset underlying this study is that despite the sample of test scores being limited to the 2017–2023 timespan, English learner students' first taking of the ACCESS test can be traced back even prior to 2017 (for states that were part of the Consortium since WIDA's inception in 2006.) [37] In other words, while the reported measurements of EL proficiency are first available in 2017, there is additional data on students' time that they have been identified as EL, insofar they have taken ACCESS before 2017.

[37] Otherwise, all EL students Time as EL would be (close to) 0 in 2017, and 1 in 2018, and so on. This would render TEL ineffective as a predictor, as it would be collinear with time, grade, and other temporal fixed effects.

However, while I can measure very precisely – in days and even minutes – the time elapsed since the students' recorded ACCESS assessment in a particular school year, due to differing academic calendars resulting in an uneven start of instruction across many schools, districts, and states, it is not feasible to accurately calculate the duration between the students' enrollment in a given school/ language support program and their test date. Furthermore, not all English learner students enrolling in K-12 schools each year are provided with a secondary language support program immediately upon identification as English Learners. Given these limitations, for each student taking their first online ACCESS assessment, the *Time as EL* variable is calculated assuming a uniform school and program start date of September 1st. So, for example, if a hypothetical student's first ever ACCESS test was recorded on May 1st, 2022, the *Time as EL* for that student (in the school year 2022) would be calculated as the interval between the latter date and September 1st, equaling to 242 days (or 0.66 years). Further, if this same student took the ACCESS assessment again next year in 2023, their *Time as EL* would equal to 1.66 years (assuming another test date of May 1st). Indirectly, *Time as EL* serves as a proxy variable measuring, in days, the amount of English language development between test administrations. [38] I also include a quadratic term for this variable to explore non-linear trends in students' language development across time, and to capture any diminishing returns from extended stay in English language support programs and EL status. Table 3.6 provides the average number of years students were identified as English learner, by grade and by year. For example, the very first cell of the table implies

[38] While this variable is captured with some measurement error, some of this would presumably be absorbed by school, district, and state-level random effects that are included in the regression analyses.

that the sample of ELs who were enrolled in the first grade in the school year 2017 reported an average time in program of 1.3 years.[39] Following up the results across years, this estimate does not change substantially from 2017 to 2023 for ELs enrolled taking the test in the first grade. This is different, however, for ELs enrolled in higher grades.

Table 3.6. Average time as English learner, in years, by grade and school year.

Grade/Year	2017	2018	2019	2020	*2021**	2022	2023	Grade Average
1	1.3	1.3	1.3	1.3	1.4	1.1	1.3	1.3
2	2.1	2.1	2.2	2.1	2.3	2.1	1.9	2.1
3	2.8	2.9	2.9	2.9	3.1	3.0	2.8	2.9
4	3.2	3.6	3.6	3.6	3.8	3.7	3.6	3.6
5	3.4	3.8	4.2	4.2	4.4	4.4	4.1	4.1
6	3.5	3.9	4.4	4.6	4.9	4.8	4.6	4.5
7	3.6	4.0	4.4	4.8	5.3	5.3	5.1	4.7
8	3.8	4.1	4.5	4.8	5.5	5.7	5.5	4.9
9	3.4	3.8	4.2	4.2	5.2	5.1	5.2	4.5
10	3.5	4.0	4.4	4.7	5.0	5.5	5.4	4.7
11	3.4	4.0	4.6	5.0	5.4	5.5	5.6	4.9
12	3.7	4.1	4.8	5.4	5.7	6.0	5.9	5.2
Year Average	2.9	3.2	3.4	3.6	3.9	3.9	3.9	3.6

Importantly, these descriptive estimates of *TEL* are aggregated (by grade) across very diverse samples of students across vastly different geographies. Furthermore, they are affected by complex and unobserved (at this scale) factors such as student mobility, dropout, and reclassification rates, that in turn vary by ELs' grade and grade-level cluster, among other factors. However, recalling the notation of the conditional formatting where green shading indicate higher numbers, the aggregate trends shown in Table 3.6 suggest that EL students are staying in language support programs for increasingly extended

[39] This estimate is different from 1.0 years to the extent that there are grade 1 English learners who have been retained a grade.

periods. This overall trend is especially evident in post-pandemic years and higher grade-levels. For example, the green shading for the post-pandemic years and middle and high school grades indicating a range from 4.9 to 6 years of average *Time as EL*, implies that more than half of ELs in these grades and years would be captured by the Long-term label (because they have been in program for 5 years, or longer).

Newcomer and Long-term English Learners

Two additional variables capturing individual-level temporal effects – *Newcomer EL* and *Long-term EL* – are directly calculated from the *Time as EL* variable. Given the multitude of contextual issues in serving the unique needs of these at-risk student populations, a large body of literature discusses the complexities of providing an effective and equitable education to these students. These two variables are included in the regression analyses to: a) further improve the precision of the temporal parameter estimates (since *Time as EL* is not measured precisely, especially in the first year of identification, due to the assumption of the uniform school/EL instruction start of September 1st), b) control for any additional effects from taking the ACCESS test for the first time ever for newcomer students, and c) capture any further (in addition to the quadratic term) diminishing returns from extended *Time as EL* - after 5 years, as stipulated by federal non-regulatory guidance- for "long-term" English Learners.[40] In regression models, the coefficients on both of these variables are expected to have negative signs, capturing the average lower proficiency of EL students who could be identified as either *Newcomers* or *LTELs*. Tables 3.7 and 3.8 provide the proportions of

[40] According to federal non-regulatory guidance, this reporting requirement does not establish a universal definition; rather, "the reporting requirement under ESEA Section 3121(a)(6) may be instructive in determining which ELs served under Title III are long-term ELs" (U.S. Department of Education, 2016, p. 38).

Newcomer and *Long-term EL* students by grade and year relative to the overall student population in that grade and year.

Table 3.7. Proportions of Newcomer English students by grade and year.

Grade / Year	2017	2018	2019	2020	*2021**	2022	2023	Grade Average
1	16	12	13	14	12	24	16	**17**
2	13	10	10	11	7	12	13	**11**
3	11	9	9	10	6	10	12	**10**
4	15	9	9	10	6	10	11	**10**
5	18	12	10	11	7	11	13	**11**
6	19	15	13	13	8	12	14	**13**
7	19	16	14	14	8	13	14	**14**
8	19	15	14	15	8	13	15	**14**
9	30	22	23	27	13	25	24	**24**
10	21	15	15	16	10	18	16	**16**
11	20	15	12	13	9	15	16	**14**
12	16	12	10	9	7	11	12	**11**
Year Average	**17**	**13**	**12**	**13**	**8**	**16**	**15**	**13**

But for the large dip in the 2021 during-COVID school year, the rates of Newcomer students have been quite stable within grades.[41] The largest proportion of Newcomers is generally observed in Grade 9, likely due to demographic trends and relatively more students being identified as ELs for the first time in high school. But for a large increase in the first grade in 2022 owing to the large number of not tested students in 2021, pre- and post-pandemic trends in Newcomer rates are not substantially different, with averages in 2022 and 2023 being slightly higher than those reported across 2018-2020.

[41] This may point to a negative *Newcomer* selection bias, in that some states and districts may have been purposeful in intentionally assessing the relatively higher proficiency subgroups in the 2021 during-COVID school year.

Table 3.8. Proportions of Long-term English Learners by grade and year.

Grade / Year	2017	2018	2019	2020	*2021**	2022	2023	Average
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	1	1	1	0	1	0	0	0
4	4	5	5	5	6	5	4	5
5	32	47	56	60	62	65	60	56
6	30	42	53	60	62	64	61	55
7	29	39	48	55	61	63	61	52
8	28	38	44	51	57	63	59	50
9	25	32	39	40	49	51	51	42
10	24	32	39	45	45	52	50	42
11	20	31	39	46	48	50	51	42
12	21	30	39	48	50	54	51	44

Parallel to the trends presented in Table 3.6 on average Time as EL, the proportions of students who would fall under the category long-term EL (LTEL) increased over time. Interestingly, LTEL rates are relatively higher in middle school (ranging from 50-56% post-COVID), compared to the rates in high school (40-50%).

For the purposes of the analytic strategy in this study, Tables 3.7 and 3.8 highlight that there is significant variation in these variables across grades. This provides further support for examining their relationship to EL proficiency in the context of regression models that include grade fixed-effects, among other variables that are related to EL proficiency.

<u>Age / SLIFE</u>

Many English learner students enrolled in US schools arrive with *limited or interrupted formal education* (SLIFE). These students are children in grades four to 12 who have experiences disruptions to their educations in their native countries and/or the United States, and/or are unfamiliar with the culture of the schooling (U.S. Department of

Education Newcomer Toolkit, 2017) The high mobility of the EL population, along with the immigrant backgrounds, and migrant statuses of many English learners are some of the factors elevating the importance of measuring any potential differences in the proficiency and language development process for these language learners that are at a further disadvantage, as compared to their peers who typically transition from elementary to middle to high schools in a more seamless and uninterrupted fashion. Further, students' educational background is important to consider because it shapes their academic journeys in important ways. Students' prior knowledge, experiences with schooling, and documented schoolwork all impact their learning trajectories in US schools, and their college and career readiness (DeCapua et al., 2009; New York State Education Department, 2011; Calderón, 2008; Short et al., 2012). Studies also show that students' home language proficiency is positively associated with English language acquisition and other academic outcomes (August et al., 2009; Calderon et al., 2011; Walqui, 2000).

Data on the students' age (measured at the time of taking the assessment) provides a simple, yet convenient way to quantify the potential impact of interruptions of formal education on ELs' average language proficiency. To this end I calculate the difference between the student's reported age (at the time of the test) and the average age of their cohort – i.e., the grade they were enrolled in when they took the assessment. This difference, measured in days and rounded to the year, is another innovation of this study, and serves as a proxy variable aimed at quantifying the hypothesized detrimental impact of interruptions to formal education (*SLIFE*).[42]

---

[42] While it is also important to examine how the students' age affects their language learning including both Age and SLIFE, expectedly, results in collinearities in the regression models.

Gender

Findings from the general student population show that female students generally outperform their male peers on some measures of academic achievement and standardized assessments such as verbal and reading tests, while the converse is reported for math and science tests (Hyde & Linn,1988; O'Dae et al, 2018; Quinn & Cooc, 2015). More pertinent to the content and test performance requirements of ACCESS, Balart & Oosterveen (2019) report that females show more sustained performance during test-taking than males in the cognitive domains where they perform both relatively better (reading) and relatively worse (math-science). Studies also show that female students have more self-discipline (Duckworth & Seligman, 2006), report fewer behavioral problems (Jacob, 2002), and display more developed attitudes towards learning (Cornwell et al, 2013). Further, while there is some evidence that female students outperform male students in second language acquisition (van der Silk et al., 2015) none of these studies offer evidence in the EL context.

Gender in the ACCESS dataset is reported as either *Male* or *Female*. *Female* students make up about 56% percent of the sample. For a very small proportion of students – under 0.5% annually – the gender variable is reported as *Missing*. Because it is impossible to ascertain whether the missing observations reflect data quality issues or that the students intentionally did not ascribe to either gender category, these observations are excluded from the analysis. [43]

[43] An alternative approach is taken for the race variable, which also reports many "Missing" observations. However, because Race is central to the analysis, and is not reported for many more students compared to that for Gender, these observations are treated differently, as explained in the section on Ethnicity and Race.

IEP Status

In the ACCESS longitudinal dataset, English Learners' disability status is measured by the *IEP* (Individualized Education Plan) variable. Disability status is one of the important predictors of students' performance on the ACCESS assessment and has been widely discussed in recent literature. Studies show large and persistent disparities between the outcomes of ELs with and without IEPs. Dual-identified students tend to report lower average performance on standardized content tests and reclassify at much lower rates relative to ELs without IEPs (Sahakyan & Poole, 2022; Shin, 2020). These disparities are further visible in the higher numbers and proportions of ELs with IEPs who fall under the LTEL category (Sahakyan & Ryan, 2018; Slama, et al., 2017; Kieffer & Parker, 2016).

Annually about 12% of WIDA's EL population is identified with an *IEP (Table 3.9).* Many English learners with disabilities and are entitled to, and oftentimes do receive appropriate accommodations during the ACCESS assessment that are intended to enable these students' taking of all four individual language domains of reading, speaking, listening, and writing.[44] However, such accommodations are always imperfect, and can only partially counterbalance for the difficulties and disadvantages that these students face compared to their non-disabled peers. Further, research indicates that there is wide variability in how accessibility and accommodations are defined, interpreted and offered across states (Christensen et al. 2018; Kim et al., 2019; Shafer Wilner & Monroe, 2016;

[44] Typically, depending on the type of disability, English Learners with a documented IEP are provided accommodations during testing. There are currently 16 types of accommodations offered to ELs with disabilities on ACCESS. Those ELs who have more severe disabilities and are therefore taking the Alternate ACCESS assessment (a decision made locally in schools by the IEP team), or those who have not completed all four domains of the ACCESS assessment, are not included in the present study.

). Except for these disability-specific accommodations no other considerations are made in calculating overall composite proficiency scores for these students, having them aspire to the same high academic standards that their non-disabled EL peers face. Importantly, all English learners with IEPs (or "dually-identified" students as referred to in the literature) that are included in the study have valid overall composite scale scores, and therefore, with or without accommodations, have successfully completed their assessments in all four language domains of reading, speaking, listening and writing, while there are many students with IEPs who are unable to do so, despite the (potentially) provided accommodations during their assessment.[45] Therefore, the uncovered disparities for this subgroup should be treated as underestimates.

Migrant Status

Approximately 1.3% of WIDA's EL students are also identified as *Migrant*. According to the National Center for Education Statistics, *Migrants* are migratory workers, or the children of migratory workers, who relocate in order to obtain seasonal or temporary employment in agriculture or fishing.[46] The educational disruptions that result from multiple moves and irregular attendance are some of the factors diminishing these students' chances for academic success.[47] And especially relevant for the English learner

---

[45] Since 2019 WIDA has been providing technical assistance and tools for its member states to generate *Imputed* Overall Composite Scale Scores for eligible EL students with IEPs who were unable to take, and thus are missing one or two of the individual language domains (Porter et at., 2021). One of these methods, referred to as the "Reweighting method", designed by the author of this study in 2020, and currently most widely used in WIDA states, redistributes the weight(s) of the missing domain(s) to the non-missing ones, thereby enabling an alternate measure of overall composite proficiency for these students that can be used to make EL reclassification decisions by educators in WIDA states. https://wida.wisc.edu/sites/default/files/resource/technical-report-generating-imputed-overall-composite-scale-scores-english-learners-disabilities.pdf

[46] https://nces.ed.gov/surveys/frss/publications/2000061/index.asp?sectionid=2

[47] Recognizing the unique needs of migrant students, the Migrant Education-Basic Grant Program (MEP) was legislated in 1966 as an amendment to Title I of the Elementary and Secondary Education Act (ESSA). Following the

population under enquiry, *Migrant* ELs' academic difficulties are further compounded by language barriers, poverty, and unique health problems putting them at a further disadvantage (DiCerbo, 2001; Umansky et al, 2018; Shafer, 2001). Examining and quantifying disparities for this student population is the first step in ensuring a more equitable and quality education for these students.

LIEP Waivers

A previously unexplored subgroup (in a large-scale quantitative context), about 1.1% of ELs in the analytic sample are identified with an *LIEP waiver*. A waiver from, or refusal of Language Instructional Educational Programs (LIEP) and/or services indicates an informed, voluntary decision by a parent of an identified English learner to not have the child placed in any specialized English language development service or instructional program. This waiver of a student from being placed in a specific LIEP does not also waive the federally mandated requirement of annual language assessment of identified English learners, so ACCESS overall composite scale scores are available for students with such LIEP waivers. Notably, the sign of the coefficient for *LIEP Waiver*, signaling whether these ELs over-, or underperform in ACCESS performance as compared to those ELs who are enrolled in supplementary language services is an empirical issue, depending on the quality of these students' out-of-school academic supports, and further resting on the accuracy and appropriateness of the decision of the parents to waive school-provided services.

---

reauthorization of MEP in 1994, the program currently operates under the authority of Title I, Part C of the Improving America's Schools Act (IASA) of 1994 to provide formula grants to states for the provision of supplemental education and support services for migrant children (U.S. Congress 1994).

Ethnicity and Race

The conceptual model presented in Figure 2 lists all the individual-level variables and some intersections that were included in the empirical analysis. Two of these variables reporting ELs' Ethnicity and Race define the focal subgroups of research in this study. For example, reiterating the importance of overlapping identities and the notion of intersectionality, Smiley et al (2023) note that race is an important consideration among Latinx populations, particularly surrounding "issues of colorism", and reference Quiros and Dawson (2013) and Ribando (2007) to assert the importance of terms such as "Afro-Latino" that are used to highlight racial identities and indicate the legacy of having Indigenous, European, and African ancestry (p. 1624).

Further, within the ACCESS longitudinal dataset English learners are identified as either *Hispanic* or *not Hispanic* in the ethnicity field, and as either *Asian*, *Black/African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander or White*, in the race field. [48] Recognizing that ethnic/racial labels such as *Hispanic*, *Asian*, or *Native American*, among other categories may provide a heuristic category for understanding the intersectionality of these categories, it is important to note that this may also mask within group heterogeneity (López et al., 2018). And while terms like *Latina*/o/x may be more reflective of the current discourse, I use the terms 'Hispanic,' 'Asian,' etc. because this is how students were identified in the ACCESS data by test administrators and state and district stall.

[48] The Office of Management and Budget announced on 3/38/2024 that the U.S. government would revise how it categorizes race and ethnicity. The categories now include a "Hispanic or Latino" box that appears under a question that asks: "What is your race and/or ethnicity?" Going forward, participants in federal surveys will be presented with at least seven "race and/or ethnicity" categories, along with instructions that say: "Select all that apply."

Importantly, each year for a substantial number of ELs identified both as *Hispanic* and as *not Hispanic*, the race variable is not reported (or equivalently, for our purposes is reported as *Missing*). Rather than excluding these hundreds of thousands of students with otherwise valid assessment and demographic data, I preserve these observations by applying an interaction of ethnicity and race. Thereby, faithful to the tenets of the theoretical framework of *Intersectionality* that calls to consider multiply-intersecting and overlapping student identities, the study interrogates potential disparities, and the impact of the pandemic on thereof across subgroups that are not exclusively defined either by the students' race <u>or</u> ethnicity. Instead, leveraging the large samples of students available for analysis, the interaction of race <u>and</u> ethnicity juxtaposes outcomes for over a dozen distinct ethno-racial intersectional EL subgroups (e.g., *Asian Hispanic* and *Asian non-Hispanic, or White Hispanic* and *White non-Hispanic, or Black non-Hispanic* and *Asian non-Hispanic).* Admittedly, these average differences are estimated in reference to a subgroup of students for whom there is no discernable ethno-racial identification (i.e., *not Hispanic*, *no Race reported*). Despite this, these estimates are dually helpful in a) generating a ranking of students' proficiency by ethno-racial categories relative to this incognizable subgroup that also happens to report the lowest average proficiency, and b) recovering reliable and more precise estimates for average disparities by *Hispanic* ethnicity, for each of the reported races.

Figure 3.4 provides the overall distribution of EL students across the WIDA Consortium grouped by this ethno-racial intersection. Notably, due to differences in students' racial identification across years, the reported average percentages are approximate. Faithful to the lens of *Intersectionality* highlighting the socially-constructed, intertwined, and dynamic nature of student categories, no additional data imputation is attempted to "fix" such changes in students' racial (and other) identification. Instead, the data is "taken as reported" relying on "fuzzy set logic" stemming from the *Intersectionality* lens, as described in the analytic strategy section (Hancock, 2013 & 2007; Russell, 2023). Further following this logic and faithful to the framework of *Intersectionality* and its focus on inclusion (rather than applying imputation techniques and grouping them in researcher-assigned categories or excluding these data altogether) a small proportion of EL students who report multiple racial categories in the race field (within a single year), are grouped into a new category, labeled Mixed/Multiple Races.[49]

Figure 3.4. Ethno-racial intersectionality of WIDA's English Learners: 2017-2023 average

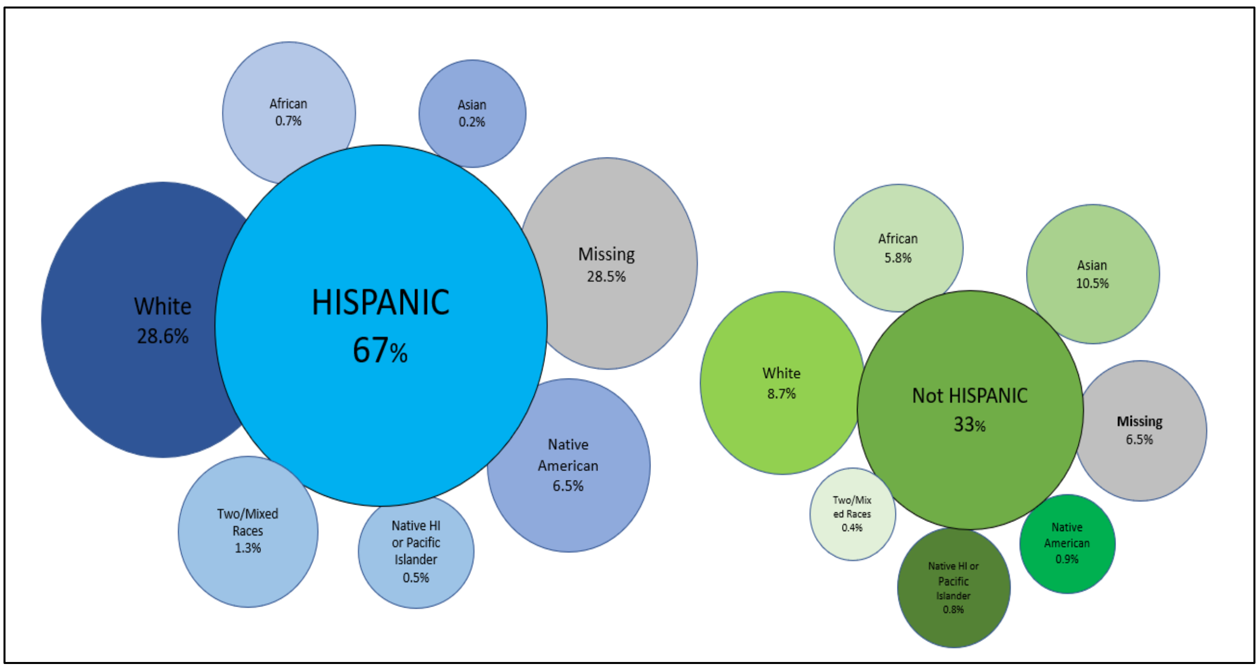


Table 3.9 presets the demographic composition of the sample for the individual-level

[49] More detailed and granular intersections are theoretically possible. For example, Pewritt (2004) mentions over 100 potential ethno-racial subgroups if a "mark one or more" approach is permitted (in Cornell & Hartmann; 2007).

variables included in the analysis, including those on students' reported race and ethnicity. In Table 3.9 the latter are presented separately, while Table 3.10 provides ELs' demographic composition along intersectional, i.e. ethno-racial categories.

Table 3.9: Demographic composition of the analytic sample.

	Student Demographics / Identities	Overall		Between		Within
		#	%	#	%	%
Gender	Male	5,293,875	54.7	1,822,611	53.7	99.8
	Female	4,390,017	45.3	1,577,540	46.5	99.7
	Total	**9,683,892**	**100.0**	**3,400,151**	**100.2**	**99.8**
IEP	No	8,499,813	87.8	3,156,090	93.1	97.0
	Yes	1,184,079	12.2	415,388	12.3	79.4
	Total	**9,683,892**	**100.0**	**3,571,478**	**105.3**	**95.0**
Migrant	No	9,613,412	99.3	3,377,322	99.6	99.7
	Yes	70,480	0.7	42,774	1.3	61.1
	Total	**9,683,892**	**100.0**	**3,420,096**	**100.8**	**99.2**
LIEP	No	9,619,091	99.3	3,378,820	99.6	99.7
	Yes	64,801	0.7	39,493	1.2	59.1
	Total	**9,683,892**	**100.0**	**3,418,313**	**100.8**	**99.2**
Ethnicity	Hispanic	3,043,008	31.4	1,271,129	37.5	93.3
	not Hispanic	6,640,884	68.6	2,253,391	66.4	97.9
	Total	**9,683,892**	**100.0**	**3,524,520**	**103.9**	**96.2**
Race	Asian	1,064,994	11.0	440,639	13.0	97.1
	African/Black	632,774	6.5	242,908	7.2	95.3
	Mixed/Multiple Races	163,766	1.7	74,528	2.2	77.6
	Native American or Alaskan	733,581	7.6	276,398	8.2	86.7
	Pacific Islander or Hawaiian	139,423	1.4	53,488	1.6	87.7
	White	3,763,989	38.9	1,426,375	42.1	92.1
	No Race	3,185,365	32.9	1,220,721	36.0	88.0
	Total	**9,683,892**	**100.0**	**3,735,057**	**110.1**	**90.8**

Table 3.9 presents the overall, between and within variations in frequency of identification for the subgroups defined by the demographic variables. For example, the first row of Table 3.9 shows that overall, 54.7 percent, or 5,293,875 of the total number of

observations in the sample (N = 9,683,892) were identified as *Male*. Examining between-student variation, 53.7 percent, or a total of 1,822,611 unique students were identified as *Male*. The 'within' variation presented in the last column, shows the within-group across-time consistency in demographic identification. Therefore, according to the last estimate reported in the row, for 99.8% of the unique student observations identifying *Male* ELs, the identification stayed constant across years. The reported total percent of between-variation for the *gender* variable, estimated at 100.2 percent shows that across 2017-2023, for 0.2% of the observations the gender recorded changed from *Male* to *Female*, or vice versa. The consistency of within-subgroup identification is high for most demographic variables (90.8% overall) and varies from about 60% for *Migrant* and *LIEP (Waivered)* ELs to close to 100% for gender identification.

Among the demographic variables the over-identification (from counting all identifications, regardless of across-year changes) for the race variables is the highest, and shows that the recorded race for about 10% of EL students changed within the seven year timespan of the study.[50] Table 3.9 also shows that the largest racial subgroup, or 40% of the total observations, is *White*, followed by the *No Race* category at about 35%. The next three largest subgroups are *Asian*, *Native American or Alaskan*, and *African/Black ELs*, representing 11%, 7.6%, and 6.5% of the total observations. *Pacific Island*er ELs represent the smallest subgroup, estimated at about 1.4% of total observations.

[50] Notably, the highest within-group consistency across time is reported by Asian and Black/African ELs, estimated at 97 and 95%, respectively.

These race-only based comparisons neglect the fact that many of the students are also assigned an ethnicity membership of being either Hispanic or not Hispanic. Conversely, the ethnicity-only based comparisons, showing that about 2/3s of the analytic sample reported Hispanic ethnicity, fail to consider students' further racial identification. Thus, for a more accurate, *Intersectional* representation of students' identities, Table 3.10 presents the ethno-racial composition of the analytic sample.

Table 3.10: Ethno-racial composition of the analytic sample.

	Ethno-racial Subgroups	Overall		Between		Within
		N	%	N	%	%
Ethnicity AND Race	Asian Not Hispanic	1,052,679	10.9	435,642	12.8	97.2
	Asian Hispanic	12,315	0.1	5,552	0.2	80.7
	African/Black Not Hispanic	583,819	6.0	221,746	6.5	96.0
	African/Black Hispanic	48,955	0.5	22,233	0.7	83.0
	Mixed/Multiple Races Not Hispanic	36,124	0.4	19,072	0.6	78.4
	Mixed/Multiple Races Hispanic	127,642	1.3	56,083	1.7	76.5
	Native American or Alaskan Not Hispanic	85,972	0.9	38,508	1.1	83.2
	Native American or Alaskan Hispanic	647,609	6.7	240,111	7.1	86.5
	Pacific Islander or Hawaiian Not Hispanic	85,396	0.9	31,192	0.9	90.8
	Pacific Islander or Hawaiian Hispanic	54,027	0.6	22,785	0.7	81.7
	White Not Hispanic	877,514	9.1	374,097	11.0	89.0
	White Hispanic	2,886,475	29.8	1,082,682	31.9	90.6
	Not Hispanic (No Race)*	321,504	3.3	234,370	6.9	80.4
	Hispanic	2,863,861	29.6	1,025,044	30.2	91.0
	Total	**9,683,892**	**100.0**	**3,809,117**	**112.3**	**89.1**

Hispanic Interactions

Further guided by the lens of *Intersectionality*, interactions of Hispanic ethnicity with students' *gender*, *IEP*, *migrant*, *waiver*, *newcomer*, and *long-term EL* status are also examined to identify any potential differences in proficiency for these intersectional subgroups. While many other potential intersections, including those with more than two overlapping identities could be considered, given the above discussed increasing concerns with the declining academic outcomes of this fast-growing population, in this work I focus on Hispanic ethnicity-centered interactions leaving other and higher-level intersections of student identities for future research. As advised by Misra (2021) et al.:

"*Researchers should consider which intersections matter most for the research question being posed, focusing on the intersections that seem most salient based on the research focus. No one project can cover every base; yet, they can be designed creatively to consider how simple additive categories may not fully uncover the social processes of interest." (p.5).*

Concluding the section describing the underlying dataset, Table 3.11 lists all the variables included in the analysis. Variables identifying *Hispanic* ethnicity and intersections with *Hispanic* identification are highlighted in light blue.

Table 3.11: List and description of variables included in regression models.

Variable Names	Type of variable	Description
Overall Composite Scale Scores (CSS)	interval	Dependent variable: English Learners' proficiency
COVID-19	binary	Impact of COVID-19 on CSS: = 1 when SY = 2021, 2022, or 2023
Grade	categorical	Grade fixed-effects, ommited from tables (reported in Appendix A)
Time as EL (in Years)	interval	Time since first ACCESS test, in years
TEL Squared	interval	Time since first ACCESS test squared, in years
Newcomer	binary	Newcomer student, = 1 when TEL < 1
Long-term EL	binary	Long-term EL student, = 1 when TEL > 5
SLIFE	interval	Deviation from cohort average age, in years
Female	binary	ELs identified as female
IEP	binary	ELs identified with an IEP
Migrant	binary	ELs identified as Migrants
LIEP Waiver	binary	ELs with Waiver from LIEP services
Asian not Hispanic	binary	Asian & not Hispanic ELs
Asian Hispanic	binary	Asian & Hispanic ELs
Black / African not Hispanic	binary	Black/African & not Hispanic ELs
Black/African Hispanic	binary	Black/African & Hispanic ELs
Mixed Multiple Races not Hispanic	binary	Mixed/Multiple Races & not Hispanic ELs
Mixed / Multiple Races Hispanic	binary	Mixed/Multiple Races & Hispanic ELs
Native American or Alaskan not Hispanic	binary	Native American or Alaskan & not Hispanic ELs
Native American or Alaskan Hispanic	binary	Native American or Alaskan & Hispanic ELs
Pacific Islander or Nat HI not Hispanic	binary	Pacific Islander or Native Hawaiian & not Hispanic ELs
Pacific Islander or Nat HI Hispanic	binary	Pacific Islander or Native Hawaiian & Hispanic ELs
White not Hispanic	binary	White & not Hispanic ELs
White Hispanic	binary	White & Hispanic ELs
Not Hispanic (No Race)	binary	No Race & not Hispanic ELs
Hispanic (No Race)	binary	No Race & Hispanic ELs, baseline for ethno-racial categories
Hispanic & Newcomer	binary	Hispanic & Newcomer ELs
Hispanic & LTEL	binary	Hispanic & LTEL ELs
Hispanic & Female	binary	Hispanic & Female ELs
Hispanic & IEP	binary	Hispanic & IEP ELs
Hispanic & Migrant	binary	Hispanic & Migrant ELs
Hispanic & Waiver	binary	Hispanic & Waiver ELs
School year (2017-2023)	integer	School year, from 2017 to 2023
State (34)	string/id	WIDA States: 34
District (7,619)	string/id	Unique District identifiers: 7,619
School (43,183)	string/id	Unique School identifiers: 43,183
Student (3,391,969)	string/id	Unique Student identifiers; 3,391,969

Analytic Strategy

The large number of student-level demographic variables and their interactions, along with the potentially important institutional context of schools, districts, and states depicted in the conceptual model presented in Figure 2 require detailed analyses and comparisons of average proficiency outcomes across time (pre- and post-COVID-19 periods) and multiply-categorized EL student subgroups. Analogous comparisons were provided in the Data section, with the (by-grade) juxtaposition of average EL proficiency outcomes pre- and post-COVID-19 in Table 3.3, or that by students' Ethnicity category, in Table 3.4. These descriptive comparisons of average proficiency leveraged the large samples of language learners taking the ACCESS online assessment to provide preliminary evidence corroborating the hypothesized large and differential impact of the pandemic on ELs and various EL subgroups. However, such high-level comparisons of average outcomes, while informative, could also potentially distort estimates and mask important differences in outcomes that may be related to important observable (and unobservable) individual- and aggregate-level factors. This concern is especially relevant for English Learners, whose educational opportunities and academic experiences are affected by inherent and interrelated factors such as diverse, dynamic, and mobile student populations, and tremendously different local contexts shaping academic experiences and outcomes. Therefore, a more rigorous analysis is warranted to examine, identify, quantify, and "take into account" (at least) for some of these salient factors that demarcate disparities for many student subgroups.

<u>Intersectional Approaches to (Examine) Complexity</u>

McCall (2005) further differentiates between three approaches to intersectional research, grouping them as *anti-categorical*, *intra-categorical*, and *inter-categorical*. The first approach is based on a methodology that rejects rigid analytical categories and aims to deconstruct them. The third approach, labeled as inter-categorical, requires researchers to provisionally adopt existing categories to document relationships of inequality among social groups and changing configurations of inequity among multiple and conflicting dimensions. Further, McCall (2005) situates the second, intracategorical approach in the middle of the continuum between first and third approaches, “as the first one rejects the rigidity of categories themselves, while the third uses them strategically”, p. 1774. The intracategorical approach, akin to the anti-categorical approach, while interrogating the process of defining (categorical) boundaries itself, focuses on particular social groups at neglected points of intersection in order to reveal the complexity of lived experience within such groups. While McCall admits that (a) not all research on *Intersectionality* can be classified into one of the three approaches, and (b) that some intersectional research crosses the boundaries of this continuum, and (c) that it is easy to misclassify intersectional research, the present study incorporates elements from all of these different approaches. Recognizing the socially-constructed, institutionally-affected, and dynamically-changing nature of many of the student-level categories – and especially that of an English Learner located at the center of the inquiry – the stable and consistent relationships and trends in outcomes that different student subgroups (e.g. Female EL students, Hispanic EL students, Black EL students) located within this larger social group (ELs) exhibit must also be emphasized. Simultaneously, the study is interested in interrogating dynamic and overlapping intersections of disadvantage, i.e. the academic experiences of EL students representing otherwise-neglected identities in the intersection of ethnicity and race, or ethnicity and gender, or ethnicity and disability status.

Regression Analysis as an Intersectional Tool

Intersectional research aims to capture the multidimensionality of students' lived experiences within the institutional contexts that shape those experiences, and continue to propagate these inequities (Russell, 2023). Quantitative modeling of *Intersectionality* has been used in the field of education by researchers trying to disentangle, identify, and assess the complex relationships and interactions between multiply-overlapping marginalized identities, structural inequities, and systems of oppression that have spawned these inequalities (Bauer et al., 2021; Le et al., 2024; Sahakyan & Poole, 2022; Warner, 2008; Weldon, 2006). For example, in their seminal study *"Making the Invisible Visible: Critical Race Theory and Intersectionality for Contextualizing Race-Gender-Class 'Achievement Gaps' in Higher Education"*, López et al. (2018) highlight the importance of addressing inherent structures of settler colonialism and the interplay of race, gender, and class inequalities in understanding six-year college graduation rates at a large public university in the US Southwest. By developing 20 distinct race-gender-class categories to analyze social experiences, they demonstrate how intersectional approaches can reveal hidden aspects of inequality in higher education.

Similarly, the regression analysis methods applied herein to examine differences between ethno-racial categories of student subgroups and decompose variations in students' proficiency present tensions and offer advantages for addressing these purposes. Rusell (2023) outlines three recommendations on how regression analysis methods can align closer to the objectives of *Intersectional* research. The first recommendation, stemming from the core tenets of *Intersectionality*, is to try and steer away from a discrete, dichotomous, or binary understandings of individual identity categories. Instead, under an *Intersectionality* lens, “identity more closely resembles a continuous variable that contains spaces between the traditional nominal variable”, p. 330. Given the influence that context and time can have on identity and social position researchers suggest attending to systemic variations among members of categories (Russell, 2023) and that “fuzzy set logic” can be useful in attending issues of within-group diversity in a manner that is substantively and theoretically consistent with the claims of *Intersectionality* (Hancock, 2013 & 2017; Ragin, 2008). As an example of “fuzzy set logic” Russell (2023) suggests using multiple categories of racial identification instead of just one, or assigning weights to the racial categorization depending on the frequency one identifies with a certain racial category.

There are several ways “fuzzy set logic” is implemented within this study. One example is the inclusion of variables on ethnic and racial identification without performing additional data imputation and “cleaning" to address seeming inconsistencies in ethnic or racial identification across time. Thereby, racial, and ethnic identifiers are permitted to remain somewhat fluid and dynamic across time. Meanwhile the large sample sizes underlying the analysis enable the estimation of consistent and precise parameters despite a degree of error, or “fuzziness” in the ethno-racial identification. Similar “fuzzy set logic” is applied to the rest of the demographic variables, allowing them to vary with time, just as the data presented itself.

Another important example of the application of “fuzzy set logic” is the use of the *“Not Hispanic”, No Race reported”* subgroup as the reference category for ethno-racial categories. While this baseline subgroup is not clearly defined with respect to its ethno-racial contours, its inclusion as the baseline category enables the uncovering of consistent differences in average proficiency across several subgroups based on race and especially ethnicity.

Russell’s (2023) second recommendation is closely related to the purpose of pursuing a more nuanced understanding of ethno-racial disparities and offers the application of interaction effects as a step to get closer to reflecting the compound functioning or intersections of identity and oppression (as well as advantage). Following this recommendation, in the specified regression models I examine several interaction terms: importantly those of (Hispanic) *Ethnicity* and *Race*, but also with other demographic variables. While still with some caveats, this approach offers more nuanced insights into the academic experiences and outcomes of students representing multiply overlapping identities.

Russells’ (2023) third recommendation to examine outcomes through an *Intersectional* lens is through specifying and estimating *multilevel* regression models. According to Scott & Siltanen (2017) “the conceptual underpinning of multilevel modeling is to explicitly account for the social contexts of inequality by animating context itself as a unit of analysis and source of variance”, p.380. The analytic strategy adopted in this study incorporates all of these recommendations simultaneously, by (a) applying “fuzzy set logic” in defining focal variables of interest and (b) examining the relationship between several main and interaction variables and EL proficiency across time via (c) multilevel models that account for variations in students' scores that can be sourced to school, districts, and states where their attend school and take WIDA's ACCESS English language assessment.

Regression Methods and Analytic Sample

Regression analysis methods are also appropriate in the context of the underlying analytic sample, which is unique in that it is not only tremendously large in scale (just under ten million total observations across students and time) and therefore statistically powered to support precise estimation of parameter estimates enabling comparisons of average outcomes for many previously neglected and intersectional EL student subgroups; moreover, the data captures the full population of research interest, including all English Learner students attending schools across 34 WIDA states. Issues related to unrepresentative, skewed, selected, or underpowered samples that typically plague statistical analyses of relationships in empirical education research are thus negligible.[51]

Moreover, the large scale of individual-level outcome data measured repeatedly across many years (for many ELs) provides longitudinal connections for many students (Table 3.5) and therefore more accurate examinations of English proficiency across time, while complete data on EL students' enrollment and nesting across WIDAs' many states, districts, and schools, enables analyses of variations in students' outcomes that are related to these institutional-level factors. This can be achieved by the inclusion of fixed- and random-effects in regression models to account for student-, school-, district-, and state-level factors, further informing the estimated relationships and providing more

[51] As described in the Data section, a small number observations was excluded due to missing data on focal variables of interest.

precise estimates of the impact of COVID-19 on average proficiency, EL subgroup disparities, and the impact of the pandemic on these disparities.

Regression Methods and COVID-19

Finally, from a research design perspective, the COVID-19 pandemic can be viewed as a “natural experiment”, assigning pre- and post- “treatment” groups of ELs that are affected by the pandemic under vastly different individual- and institutional-level circumstances and factors, but during similar timespans, and taking the same outcome assessment. Some students in the analytic sample took the ACCESS online assessment in pre-COVID years (2017-2020), others were tested only in post-COVID years (2021-2023), and still others took the test during both periods. All of these students’ recorded outcomes of English proficiency are included in the regression analysis to estimate the average impact of the treatment, i.e., the pandemic. The latter is estimated directly in regression models by including a COVID-19 binary variable, taking the value 1 for all student-level observations in the post-COVID-19 period. The cumulative and multi-faceted impacts of the pandemic on students, schools, districts, states, and the whole education system serving ELs, are thus measured by the coefficient of the COVID-19 parameter that quantifies pre- and post-COVID differences in average EL outcomes. Further, the differential impact of the pandemic, specific to individual- and institutional-level factors, can be estimated through the same regression model under pre- and post-COVID-19 conditions, i.e., separately for the 2017-2020 data and for the 2021-2023 data the difference between the two sets of estimated parameters can then be compared. Meanwhile, the impact of the pandemic can be estimated directly in regression models by including a COVID-19 binary variable, taking the value 1 for all the student-level observations in the post-COVID-19 period. Thus, the cumulative and multi-faceted impacts of the pandemic on students, schools, districts, states, and the whole education system serving ELs, are measured by the coefficient of the COVID-19 parameter that quantifies pre- and post-COVID differences in EL outcomes. Further, the differential impact of the pandemic, specific to individual- and institutional-level factors, can be estimated through the same regression model under pre- and post-COVID-19 conditions, i.e. separately with 2017-2020 data and 2021-2023 data, and taking the difference between the two sets of estimated parameters.[52]

Importantly, while the latter parameter estimates have a "causal flavor", i.e., could be interpreted as causal estimates of the impact of the COVID-19 pandemic on students' outcomes, the same cannot be said about the relationships between other covariates and EL proficiency. This is due to several reasons, including potentially important and omitted variables, and issues related to potential selection bias, as students' assignment to schools, districts, and states is not random. While the inclusion of multilevel random effects is aimed to reduce the sources of unobserved heterogeneity at the student-, school-, district- and state-levels, the parameter estimates implying differences in outcomes across individual- and institutional-level factors should still be interpreted as correlational and descriptive. This is not only due to the potential of omitted variables and (remaining) selection bias, but importantly because most of these variables, and especially those ascribing students' ethno-racial identification – as obviated by the tenets of *Intersectionality* – <u>should not and do not</u> have a causal impact on students' measured

[52] An equivalent approach would be to interact the COVID binary variable with all the model coefficients. However, in the context of the large samples this introduces very high computational requirements on the multilevel models.

proficiency. Regardless, such analyses can provide useful evidence of underlying relationships and inequities within systems, without asserting that those relationships are causal (Loeb at al., 2017). Identity markers are predictive of disparities due to underlying mechanisms of disadvantage, so the differences are observable; but however salient and tangible, these differences are not due to identity, but rather products and features of the education system that intervene on students' proficiency in ways connected to their identities (Poole, 2024, forthcoming).

Modeling Approach

To the best of my knowledge, this work is the first large-scale empirical examination of outcomes within the English Learner population across states, districts, schools, and time. Moreover, owing to the unique dataset at the heart of this study, it may be the first examination of student (or any) outcomes, attempting to estimate statistical dependencies in a five-level nested structure, modeling both temporal and institutional-level variation. Further, given the limited, and largely context-dependent empirical evidence on within-EL differences in academic outcomes, as well as on the potential impact of the pandemic on individual- and aggregate-level factors related to these outcomes, I consider, examine, and discuss several model specifications with increasing complexity and flexibility. The results of these interconnected models can inform the work of other researchers who are interested in conducting intersectional research using large datasets that include potentially different levels of available data and variables available for analyses.

I introduce models with increasing levels of complexity and flexibility while exploring alternate specifications that both test the sensitivity of results and further inform various relationships and dependencies between EL proficiency and its predictors. More specifically, the parameter estimates of focal variables of research interest are examined through 7 primary and 18 auxiliary regression models, as sets of individual-level variables (and their interactions) and *Student*-, *School*-, *District*-, and *State*-level effects are gradually introduced in subsequent models to explore differences and variations in students' average proficiency across various model specifications. In estimating the longitudinal and mixed-effects models that account for both the multi-level / hierarchical and repeated nature of the data to decompose variations in EL proficiency outcomes across several EL subgroups, I follow Rabe-Hesketh and Skrondal's (2021) comprehensive guide on specifying and implementing longitudinal and mixed-effects models in STATA.

All regression models assume a linear relationship between the dependent and independent variables. This gives a straightforward interpretation of the parameter estimates of the coefficients, scaling them in the dependent variables' units, i.e., as scale score changes in EL proficiency related to unit changes in the predictor variables. Since most of the predictors in the models are represented by dichotomous variables and their interactions (e.g., *Female*, or *Female* and *Hispanic*), the coefficients for these variables measure the change, in *CSS* points, in the outcome variable (EL proficiency), that is associated with membership to the subgroup as compared to a corresponding baseline category. For example, the estimated coefficient for the *Female* variable would indicate the difference, in *CSS*, between *Female* and *Male* ELs' average scores, while the coefficient for *Hispanic* and *Female* would measure, in *CSS*, any additional positive or negative differences associated with the average performance of ELs that are (additionally) located in the intersection of those identity markers.

I begin examining students' average English proficiency in a linear Ordinary Least Squares (OLS) specification, and gradually add sets of student-level demographic variables in *Models 0-2*. *Model 3* introduces a Generalized Least Squares (GLS) specification accounting for the repeated nature of observations across time. *Models 4-7* are hierarchical, or *Mixed-effects* models with random intercepts at the student-, school-, district-, and state-levels. OLS models are estimated using school-level clustered robust errors, while longitudinal and mixed-effects models impose an autoregressive (AR1) error structure (at the student-level). In auxiliary regressions, the models denoted with a '*b'* and *'c'* suffix examine specifications adding *State*- and *District*-level fixed-effects, respectively. Auxiliary models '*d'* examine the relationship between the covariates and EL proficiency using *Year* fixed-effects instead of the binary *COVID-19* variable, positioning school year 2017 as the baseline. This specification allows for further decomposing the average impact of COVID into yearly differences in average proficiency and examining annual trends in aggregate language development across time. Auxiliary models '*e'* examine the effect of "decoupling" of *ethnicity* and *race*, by including these variables into respective OLS, Longitudinal, and Mixed-effects models as separate and independent (not interacted) variables. [53,54] The last four auxiliary models '*f*', '*g'*, '*h',* and '*i'* are estimated

[53] While the Intersectionality framework suggests that this decoupling is technically incorrect, I provide these data in this work in secondary regression models acknowledging that many states and districts may not enroll sufficiency large and diverse samples enabling such multi-categorical comparisons. Further, the decoupling of race and ethnicity enables cross-model comparisons, and quantifying the extent of error in estimating ethnic disparities when students' ethno-racial intersectionality is neglected.

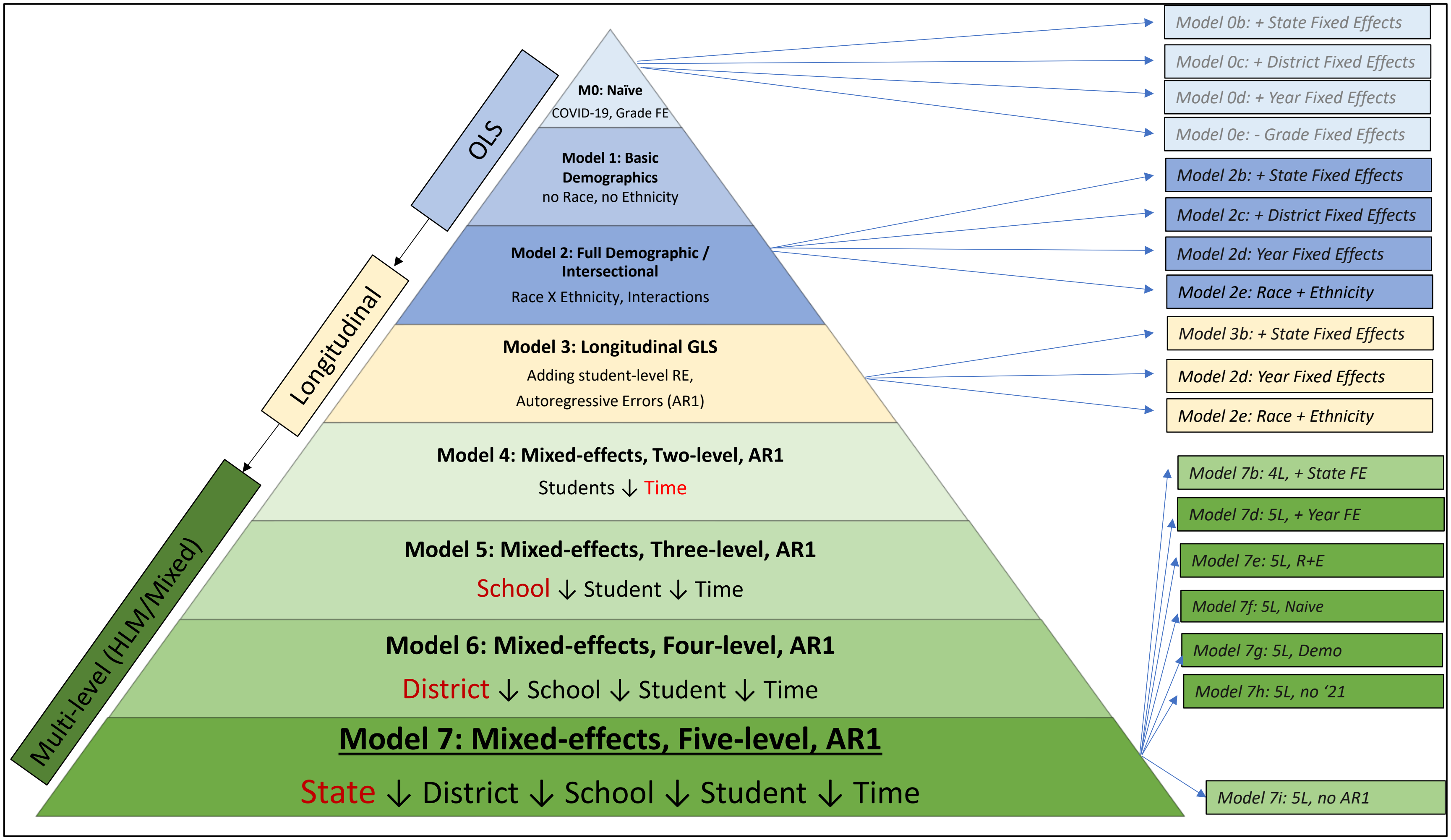
[54] In the "decoupled" models the baseline category for Ethnicity is 'not Hispanic', and for Race – ' No race provided'.

only for the final *Model 7*. *Models 7f* and *7g* examine alternate specifications with excluded sets of demographic variables, providing a comparison with baseline *Models 0, 1,* and *2* which are estimated with the same sets of independent variables, but without the inclusion of the institutional effects, i.e. students' nesting across schools, districts, and states. *Models 7h* and *7i* examine the robustness of temporal effects, by removing from the analytic sample the data from the 2021 school year, and by examining a specification that assumes no serial correlation by removing the imposed AR1 structure.[55]

For simplicity, I only present the full specifications for *Model 2, Model 3* and *Model 7*, which represent the final models using OLS, GLS (longitudinal) and Mixed-effects specifications for empirical estimation. The summary output of these regressions for the main models is given in Appendix A. Figure 3.5 presents a visual representation of the full list of estimated models.

[55] This enables the calculation of intra-cluster correlation (ICC), available only in an unconditional RE setup.

Figure 3.5: Full list of main and auxiliary regression models.



Regression Models

OLS Models

The regression analysis opens with a *Naïve Model 0*, which includes a constant and a *COVID-19 binary* variable to establish a baseline estimate for the average *impact of COVID-19* on EL proficiency. (Grade effects are included in the models by default.) *Model 0* also establishes a baseline estimate for the constant and for the conditional total residual variance, which is of research interest as institutional-level factors (i.e., higher-level random-effects) are gradually introduced into the model to adjust for additional levels of variation in students' scores.[56]

Several auxiliary regression models are estimated to examine how regression coefficients adjust to "differencing out" the state-level variation in *Model 0b* and district-level variation in *Model 0c* through *State-* and *District* fixed-effects on these baseline OLS estimates. *Model 0d* replicates the *Naïve* model using *Year* fixed-effects instead of a binary *COVID-19* variable, thereby providing a baseline estimate of annual changes in aggregate proficiency (relative to school year 2017). *Model 0e* removes the *Grade* fixed-effects for a fully unconditional (across-grade) juxtaposition of EL proficiency outcomes pre- and post-COVID-19, as well as the estimate for unconditional total variance of EL proficiency. This "Naïve" estimate of the average impact of *COVID-19* is expected to be identical to the average difference in overall composite scale scores in Table 3.3 (See Chapter 3 on methods).

---

[56] (Adjusted) R-squared is also calculated in OLS and GLS regressions for model fit comparisons.

Next, building on the *Naïve* models, in *Model 1* I add demographic variables on the students' *Time as EL (TEL)*, *a quadratic term* for the latter, and binary variables indicating ELs' status as a *Newcomer* or *Long-term EL*. Also included are an interval variable proxying the impact of interruptions in ELs' education - *SLIFE*, and binary variables identifying the students' *Gender*, *IEP* identification, *Migrant* status, and *Waiver from LIEP* services. *Model 1*, does not include focal variables of research interest – students' *Race* and *Ethnicity*. These are introduced in *Model 2*, specified in Equation 1 below, which includes both these variables, through an interaction term between students' (Hispanic) ethnicity with their reported race. *Model 2,* and all subsequent models also include *Hispanic* interaction variables, to estimate potentially disparate outcomes for *Hispanic Newcomers*, *Hispanic LTELs*, and *Hispanic* students identified as *Female*, those identified with *IEPs*, *Hispanic Migrant* ELs, and *Hispanic ELs* with *LIEP Waivers*.

(Equation 1)

$$CSS_{ij} = \beta_0 + \beta_1 \cdot \Delta Grade_{ij} + \beta_2 \cdot COVID_j + \beta_3 \cdot RaceXHispanic_{ij} + \beta_4 \cdot TEL_{ij} + \beta_5 \cdot TEL_{ij}^2 + \beta_6 \cdot Newcomer_{ij} + \beta_7 \cdot LTEL_{ij} + \beta_8 \cdot SLIFE_{ij} + \beta_9 \cdot Female_{ij} + \beta_{10} \cdot IEP_{ij} + \beta_{11} \cdot Migrant_{ij} + \beta_{12} \cdot LIEP_{ij} + \beta_{13} \cdot (Hispanic_{ij} \times Newcomer_{ij}) + \beta_{14} \cdot (Hispanic_{ij} \times LTEL_{ij}) + \beta_{15} \cdot (Hispanic_{ij} \times Female_{ij}) + \beta_{16} \cdot (Hispanic_{ij} \times IEP_{ij}) + \beta_{17} \cdot (Hispanic_{ij} \times Migrant_{ij}) + \beta_{18} \cdot (Hispanic_{ij} \times LIEP_{ij}) + \varepsilon_{ij}$$

, where $i = 1, 2, \ldots, 3{,}391{,}969$ is the subscript for individual students, $j = 1, 2, 3, 4, 5, 6, 7$ is the subscript for time/school years, β_0-β_{18} are the coefficients for the time-varying (as evident from the *j* subscript) level-1 covariates capturing various student subgroups and variables predicting EL proficiency, and ε_{ij} is assumed to have a normal distribution with a mean 0 and constant variance (this assumption is relaxed in advanced models).

As in the Naïve model, for the dual purpose of examining the stability of focal parameter estimates and for investigating the impact of controlling for aggregate-level variations in EL proficiency, secondary models *M2b, M2c,* and *M2d* are estimated, including *State*, *District*, and *Year* fixed-effects respectively. *Model 2e* investigates the impact of "decoupling" ethnicity and race, providing separate and independent coefficients for *Ethnicity* and *Race*.

Notably, the estimates from these OLS models (*M0-M2*) do not account for the fact that many of the student-level observations, both for pre- and post-COVID periods, are recorded by the same students. This is where longitudinal (and further mixed-effects models) become useful.

Longitudinal Model

Model 3 is the first specification to consider the longitudinal nature of the data, directly accounting for the repeated measurements of English proficiency across academic years by many of the same students. The longitudinal regression model specified in *Model 3* utilizes a Generalized Least Squares (GLS) approach and an autoregressive error structure (AR1) to estimate the temporal variations in EL students' scores, while controlling for the same (final) set of demographic variables included in *Model 2*.[57] More specifically, the specification in Equation 2, applied in *Model 3* uses a random-effects implementation, where unobserved between-student heterogeneity is represented by student-specific effects that are randomly varying across time, within-

[57] This model was operationalized in STATA 18 using the *xtregar, re* command, which fits cross-sectional time-series regression models when the disturbance term is first-order autoregressive. *Xtregar* accommodates unbalanced panels with unequally spaced observations and offers a GLS estimator for random-effects models. The latter is applied in the longitudinal regression specified in Equation 2. Baltagi and Liu (2020) show that this estimator produces the BLUP (best linear unbiased predictor) in unbalanced panels.

student. These types of models are more useful for investigating average relationships between the response variable and covariates, while also allowing for student-specific intercepts. Typical applications of random-effects longitudinal models investigate physical growth, or learning, where both the nature and reasons for individual-level differences in outcomes are of major interest. This contrasts with the fixed-effects implementation of longitudinal models, where every student would act as their own control, aimed at reducing student-level confounding and therefore facilitating causal inference (Rabe-Hesketh & Skrondal, 2021).

Further, the random-effects implementation of the GLS offers another key advantage as opposed to other longitudinal regression models, such as the GEE (Generalized Estimating Equations) approach (Liang and Zeger, 1986). In the GLS RE specification the autoregressive error structure is imposed directly on the error terms, while the GEE approach requires specifying a serially-correlated variance-covariance structure. This in turn means that the in the GLS RE approach no observations need to be excluded, while the GEE approach (and others) require that included students have at least two adjacent observations (for calculating appropriate serially correlated standard errors).

(Equation 2)

$$CSS_{ij} = \beta_0 + \beta_1 \cdot \Delta Grade_{ij} + \beta_2 \cdot COVID_j + \beta_3 \cdot RaceXHispanic_{ij} + \beta_4 \cdot TEL_{ij} + \beta_5 \cdot TEL_{ij}^2 + \beta_6 \cdot Newcomer_{ij} + \beta_7 \cdot LTEL_{ij} + \beta_8 \cdot SLIFE_{ij} + \beta_9 \cdot Female_{ij} + \beta_{10} \cdot IEP_{ij} + \beta_{11} \cdot Migrant_{ij} + \beta_{12} \cdot LIEP_{ij} + \beta_{13} \cdot (Hispanic_{ij} \times Newcomer_{ij}) + \beta_{14} \cdot (Hispanic_{ij} \times LTEL_{ij}) + \beta_{15} \cdot (Hispanic_{ij} \times Female_{ij}) + \beta_{16} \cdot (Hispanic_{ij} \times IEP_{ij}) + \beta_{17} \cdot (Hispanic_{ij} \times Migrant_{ij}) + \beta_{18} \cdot (Hispanic_{ij} \times LIEP_{ij}) + v_i + \varepsilon_{ij}$$

,where $i = 1, 2, \ldots, 3{,}391{,}969$ is the subscript for individual students, $j = 1, 2, 3, 4, 5, 6, 7$ is the subscript for time/school years, β_0-β_{18} are the coefficients for the time-varying (as evident from the *j* subscript) level-1 covariates capturing various student subgroups and variables predicting EL proficiency, $\varepsilon_{ij} = \rho\varepsilon_{i,j-1} + \theta_{ij}$, $|\rho| < 1$, and θ_{ij} is independent and identically distributed (i.i.d.) with mean 0 and variance σ_{ij}^2. Further, v_i are assumed to be realizations of an i.i.d. process with mean 0 and variance σ_v^2, and in the random-effects implementation of Equation 2, as described below, are assumed to be independent of X_{ij}.

The longitudinal specification in Equation 2 is preferred to the pooled OLS specification (even with robust standard errors) as OLS treats longitudinal data as repeated cross-sectional data (where samples of students are drawn independently at each occasion) and conflates within- and between-student comparisons. Between-student comparisons are susceptible to omitted-variable bias or unmeasured confounding, due to time-constant student-specific variables that are not included in the model. Within-student comparisons are free from such bias because students truly act as their own controls. Another crucial limitation of pooled OLS is that estimates of regression coefficients are no longer consistent if there is missing data and if "missingness" depends on observed responses for the same student. As illustrated in Table 3.5 this is certainly the case with the EL student population and its outcomes of EL proficiency, determining exit from EL status, and therefore affecting patterns of "missingness" in the data.

Within the longitudinal/GLS framework, an auxiliary model including *State fixed-effects (M3b)*, *Year-specific dummy variables (M3d),* and decoupled *Race* and *Ethnicity (M3e)* are also estimated. Secondary longitudinal models that include *District and School* *fixed-effects* are not feasible to estimate due to computational limitations stemming from the tremendously large number of districts (over 7,500) and schools (over 40,000).[58]

Estimation of variations in EL proficiency that are attributable to institutional-level factors through the inclusion of random-effects, on the other hand, is both feasible and useful as they can be further decomposed into random intercept and slope models with complicated variance-covariance structures. This approach enables more flexible and realistic assumptions regarding the within-cluster variability in students' proficiency. *Mixed-effects* models, which contain both *fixed*- and *random-effects,* are described next as the final family of regression models.

Mixed-effects Models

In the models presented above, several potential sources of variations in EL proficiency were considered through the inclusion of *Year*, *State,* and *District fixed-effects* in the auxiliary models. This approach has allowed for a more informed and nuanced view of the relationships between predictor and outcome variables, while gradually adjusting for the effects of any observed and unobserved heterogeneity that may be related to these aggregate-level factors. However, due to the very large number of students, schools, and districts, a fixed-effects approach of controlling for aggregate-level effects, while simple and straightforward conceptually, is impossible to implement at lower than the district-

[58] In a fixed-effects framework these school- and district-level parameters are estimated directly via (implicitly) including dummy variables in the model specification. This means that there would be tens of thousands of additional predictors added to the specifications, which makes the estimation technically infeasible.

level due to the above-discussed computational limitations.[59,60] Further, a fixed-effects approach – while technically equivalent to random effects in terms of statistical validity and estimated model parameters under large and complete samples like the one underlying this analysis – is not well suited for other reasons. Even if a supercomputer made it technically possible to estimate the model after the inclusion of tens of thousands of *District* and *School* fixed-effects, many of these effects would presumably be collinear across levels. In other words, models that include S*chool* fixed-effects would be fully saturated and would force the omission of additional *District* or *State* fixed-effects. This would preclude studying variations in student scores due to multiple levels of institutional effects both sequentially and simultaneously, which is one of the goals of this study as reflected through the lens of *Intersectionality* and the conceptual model (Figure 3.1).

Instead, to account for the multiple levels of hierarchical nesting and repeated observations across time, I specify and estimate *Mixed-effects* models which have important advantages in the context of the present study. These models provide a precise and consistent decomposition of variation in EL proficiency that can be sourced to (a) institutional factors/nesting, as captured via *State-, District-, and School-* random-effects, and (b) temporal variations, as modeled through the *Student-* random effects. Akin to OLS regressions, the relationship between student-level predictors such as individual-level

[59] For the longitudinal Model 3, the estimation will not converge with district fixed-effects due to the large number of districts, even after customizing STATA's default memory settings to include the maximum number of variables.
[60] Additionally, due to the large number of schools with small cluster sizes (number of ELs), there are likely many schools that will have to be omitted from estimation due to collinearity. This was a (smaller) issue at the district-level; even in secondary models including district fixed effects a substantial number of districts were omitted from estimation.

demographic variables and the EL Proficiency (c) is estimated though global (level 1), “fixed” coefficients (β_0-β_{18}).

Several other features of the random-effects approach and the underlying population / data generating mechanism further validate the use of the random-effects approach of modeling temporal and institutional-level variations. Given the sufficiently large number of clusters and exchangeability of cluster-level effects, and the overarching purpose of generalizing to the entire population, along with the lack of the need to evaluate cluster-specific differences (i.e. evaluate differences of EL proficiency between specific states, districts, schools, and students), the random-effects approach is also suitable in the context of the distributional features of the EL student population being examined. This is because in the random-effects approach the cluster sizes (e.g., number of ELs in a given school) don’t need to be large for a consistent estimation. Under these conditions (of potentially small number of ELs nested in specific clusters) the random-effects approach is considered superior because of “shrinkage” or “partial pooling”, which adjusts the estimates of group-level effects based on the amount of data available for each cluster and the variance both within and between groups.[61] For parameter estimation in the random-effects models the requirement is that there are a sufficient number of clusters of size 2 or more. Moreover, it does not matter if there are also clusters of size 1 (i.e., schools that report just one EL student in a given year and grade). Such singleton clusters do not provide information on the within-cluster correlation or on how the total variance is partitioned into the fixed and random components, but they do

[61] Clusters with less observations or higher variance are more influenced (shrunk) by the overall mean, while clusters with more observations and retain more of their individual characteristics instead of being pooled.

contribute to the estimation of coefficients and the total variance. Notably, because the aim it to generalize to the population of clusters (students, schools, districts, and states) and not just making inferences for the particular clusters in the data, this leads to a larger standard error for in the random-effects approach compared with the fixed-effects approach (Rabe-Hesketh & Skrondal, 2021). [62]

The final empirical model specification is a five-level random-intercepts mixed linear model with fixed level-1 covariates, with repeated observations of English proficiency of EL students nested in schools, districts, and states.

(Equation 3)

$$CSS_{ij} = \beta_0 + \beta_1 \cdot \Delta Grade_{ij} + \beta_2 \cdot COVID_j + \beta_3 \cdot RaceXHispanic_{ij} + \beta_4 \cdot TEL_{ij} + \beta_5 \cdot TEL_{ij}^2 + \beta_6 \cdot Newcomer_{ij} + \beta_7 \cdot LTEL_{ij} + \beta_8 \cdot SLIFE_{ij} + \beta_9 \cdot Female_{ij} + \beta_{10} \cdot IEP_{ij} + \beta_{11} \cdot Migrant_{ij} + \beta_{12} \cdot LIEP_{ij} + \beta_{13} \cdot (Hispanic_{ij} \times Newcomer_{ij}) + \beta_{14} \cdot (Hispanic_{ij} \times LTEL_{ij}) + \beta_{15} \cdot (Hispanic_{ij} \times Female_{ij}) + \beta_{16} \cdot (Hispanic_{ij} \times IEP_{ij}) + \beta_{17} \cdot (Hispanic_{ij} \times Migrant_{ij}) + \beta_{18} \cdot (Hispanic_{ij} \times LIEP_{ij}) + ST_j + D_{ij} + SC_{ij} + v_i + \varepsilon_{ij}$$

, where $i = 1, 2, \ldots, 3{,}391{,}969$ is the subscript for individual students, $j = 1, 2, 3, 4, 5, 6, 7$ is the subscript for time/school years, β_0-β_{18} are the coefficients for the time-varying (as evident from the *j* subscript) level-1 covariates capturing various student subgroups and variables predicting EL proficiency, $\varepsilon_{ij} = \rho\varepsilon_{i,j-1} + \theta_{ij}$, $|\rho| < 1$, and θ_{ij} is independent and identically distributed (i.i.d.) with mean 0 and variance σ_{ij}^2. Random effects ST_i, D_i, SC_i, and v_i are assumed to be realizations of an *i.i.d.* processes with mean 0 and variances σ_{ST}^2, σ_D^2, σ_{SC}^2, σ_v^2, and are assumed to be independent of X_{ij}. [63]

[62] Therefore, the estimated standard errors, which, (as the findings will show) are already very small, are likely overestimates.

[63] While it is possible to relax some of these assumptions, the computational limitations do not allow this in due to the tremendously large sample size, along with the large number of multi-level clusters.

Technical Notes and Considerations

The amount of real, extant data on students' observed proficiency and their demographic information that is being processed by complex and multilevel statistical models implemented in statistical software to build a list of estimates like that presented in Table 4.3 is truly extraordinary. However, while having access to the entire universe of ACCESS Online data enables many dimensions of this empirical examination, it also comes at a cost. For example, the very large sample of student assessment data included in the analysis limits the implementation of the more advanced multi-level models allowing even more realistic assumptions on dependencies via more flexible error-variance structures. Even "simpler" models with multiple levels of hierarchical nesting require several days, if not weeks of computer runtime in Stata 18, rendering statistical analyses very time consuming.[64]

Further, limited both by the feasibility of estimation from a computation burden standpoint, and the scope of the research questions in this work, in the decomposing of the variation to state-, district-, school- and student-levels I utilize the simplest structure of the random effects, and include only random intercepts at each level. The inclusion of random slopes in addition to random intercepts will allow examining potentially important differences in ways the pandemic has impacted states, districts, and schools, as well as differences in how students representing various ethno-racial subgroups are (under)served in various levels of public education.

[64] Model 7 required about three weeks of computer runtime for the likelihood maximization algorithm (based on Laplacian approximation of polynomials) to converge on the UW School of Education Remote Application server. A more flexible specification, for example using random slopes in addition to random intercepts, or applying a robust VC structure, has proven infeasible in a 5-level model, with this sample and under the current computing and statistical software limitations.

Finally, the mixed-effects models specified above utilize the simplest possible nesting structure for the repeated observations for students nested in schools, districts, and states. However, many students are not nested within the same school (and district) throughout their academic trajectory as a student and EL, as many ELs change schools (and sometimes districts) both as a part of regular transition from elementary to middle to high school, and as families relocate in search of better professional and academic opportunities. [65] To alleviate these concerns, a cross-nested multilevel structure would be more appropriate, to more precisely model and calculate the crossed random-effects parameters (and their precision). However, again due to computational limitations, a cross-nested structure is not feasible to implement in the context of the tremendously large sample with multiple levels of nesting. Under the assumed one-way nesting structure applied in the mixed models above, each across-year school move "resets" a student, treating the related random effect as generated by a new student.

Despite this, capitalizing on the ever-increasing computing power and sophistication of statistical software has allowed for modeling and quantifying important relationships and dependencies while accounting for the nuanced ways in which average English proficiency outcomes manifest across millions of students, vastly different geographies, and an extended period of time separated by the COVID-19 pandemic.[66] Findings on these relationships and dependencies and presented in the next chapter.

[65] Across-state moves are not tracked. When a student moves states and takes the ACCESS assessment in a new state they are counted as a new student.

[66] The analysis for this project started in late 2021, with the first wave of post-pandemic ACCESS data. At this time the analyses were performed using STATA 17, which required a higher model runtime for mixed models and would not execute some of the more complicated specifications using the full sample of data. Starting in late 2022 STATA 18 became available, allowing for more flexible model specifications, despite the longer computing times associated with the addition of recent post-COVID-19 data from 2022 and 2023.

CHAPTER FOUR: FINDINGS

Introduction

In this chapter I present the results of the empirical analyses produced by the different regression models specified and described in the previous chapter. Owing to the linear nature of all examined regression models, the estimated coefficients of all (fixed) covariates and the random-effects variance parameters are on the same scale as the dependent variable, i.e., ACCESS Online overall composite scale scores (CSS).[67] Further, the estimated standard errors for nearly all the included covariates are very small, likely due to the extremely large and complete underlying samples representing the entire population of EL students in WIDA states. This enables a more dedicated focus on the magnitude, rather than precision of the parameters and inquiries into the way parameter estimates change (or stay constant) across various model specifications.[68]

As detailed in the section on regression models in Chapter 3, *Models 0-2* examine relationships between covariates and EL proficiency in an OLS framework and set baseline estimates. *Model 3* adds the temporal dimension, examining aggregate EL proficiency using longitudinal regression methods based on a Generalized Least Squares (GLS) approach and student-level random effects. *Models 4, 5, 6,* and *7* are multilevel or *Mixed-effects* models, and add a hierarchical structure to the estimation by including *Student*-, *School*-, *District*-, and *State*-level random effects. OLS models are estimated under a school-clustered error variance structure to control for potential

[67] Variance parameters can also be interpreted in units of standard deviations (in CSS), by taking the square root of the reported estimates.
[68] Statistically not significant parameter estimates are highlighted by a gray shading and italicized text. Following the presentation in the previous chapters, I italicize model and variable names for better legibility.

heteroskedasticity, i.e., correlated variance within schools. Longitudinal and mixed-effects models use autoregressive errors to account for the serial correlation of student-level observations across time.

Following the organization of research questions presented Chapter 2, the presentation of results is organized into two sections. In the first section I describe findings on the average impact of the pandemic on EL outcomes, while quantifying (correlational) relationships between the various individual- and institutional-level factors and ELs proficiency under increasingly complex and flexible model specifications. I present the uncovered differences in outcomes between various ethno-racial and other EL subgroups, and outline how accounting for the nesting of outcomes within students (across time), schools, districts, and states affects the estimated relationships and differences. While a comparison of all coefficients across all model specifications from the three regression families in one table would be useful, due to space limitations the results from these analyses are presented in two parts: the results of OLS models are included in Table 4.1, and of GLS (longitudinal) and Mixed-effects models in Table 4.2.[69] Table 4.3 summarizes the results of the main OLS, GLS, and Mixed specifications removing the secondary models from presentation. I summarize the first section with a presentation of the results from the final model specification (Equation 3), focusing on the main *Model 7* parameters, and highlighting notable findings from the auxiliary regression models.

[69] The analysis of the coefficients was done in Microsoft Excel, which allows a simultaneous comparison of many variables and columns.

In the second section, I address the second research question and explore the impact of the pandemic on all the aforementioned factors, replicating *Model 7* specification separately for pre- (2017–2020) and post- COVID-19 (2021–2023) academic years. These results are presented in Table 4.4.

Impact of COVID-19 on EL Proficiency

Tables 4.1 and 4.2 present the complete list of models that were applied to estimate differences and dependencies across and between available individual-level variables and aggregate-level factors that could potentially impact EL proficiency outcomes across time.[70] Together, the results presented in these tables address research question 1, while controlling for individual and institutional-level factors. In Tables 4.1 and 4.2 the primary models numbered *M0-M7* (and presented separately in Table 4.3) are listed in bold, while auxiliary models examining the consistency, robustness, and sensitivity of parameters to alternate specifications are given in italic and have character suffixes (e.g. *M2b* stands for *Model 2b* – an auxiliary model to examine the impact of including *State* fixed-effects in in *Model 2*).

Each of the estimated coefficients presented in the respective cells of Tables 4.1, 4.2, and 4.3 represents a statistically significant relationship (unless italicized and grayed out), signaling an identifiable difference from a baseline subgroup, or non-trivial impact of an observed factor with respect to EL proficiency.[71] Also given in bold and highlighted in orange are three focal variable of research interest: parameter estimates on the average impact of COVID-19, the estimated average disparity between Hispanic and non-Hispanic

[70] Standard errors are omitted from the tables and are provided for the main models only in Appendix A.
[71] Statistically not significant relationships are marked by the gray shading and italic font of the cells.

students, and all the parameter estimates of covariates from the final model estimation. In Tables 4.1 and 4.2 the main model parameters are given in a larger font compared to those of secondary models. Random-effects parameters, estimates for total residual variance, R-squares (for linear models) and autoregression coefficients (in italics, for longitudinal and mixed models) are presented in the bottom part of the tables.

Table 4.1: Parameter estimates of main and auxiliary OLS specifications.

	REGRESSION FAMILY	OLS										
	Models Numbers *(DV = CSS)*	**M0**	*M0b*	*M0c*	*M0d*	*M0e*	**M1**	**M2**	*M2b*	*M2c*	*M2d*	*M2e*
	Model Name, Levels and Structure; Type of residuals and variance	**Naïve**	+ State FE	+ District FE	+ Year FE ('17 baseline)	- Grade FE	**+ Demo: no E&R**	**+ Full Demo: ExR, Interactions**	+ State FE	+ District FE	+ Year FE ('17 baseline)	Demo: E+R
	Average Impact of COVID-19	**-6.9**	-6.8	-7.0	3.6, 2.9, 1.6 -2.4, -4.8, -6.4	-4.9	**-8.0**	**-7.7**	-7.9	-8.2	2.2, 0.7, -0.5 -5.8, -7.0, -8.2	-7.8
Demographics	Time as EL (in Years)	-	X	X	X	X	9	9	9	9	9	9
	TEL Squared	-	X	X	X	X	-1	-1	-1	-1	-1	-1
	Newcomer	-	X	X	X	X	-9	-2	-2	-3	-2	-3
	Long-term EL	-	X	X	X	X	-4	-8	-8	-7	-8	-9
	SLIFE (Δ from cohort avg age)	-	X	X	X	X	-7	-6	-6	-6	-6	-6
	Female	-	X	X	X	X	5	5	5	5	5	5
	IEP	-	X	X	X	X	-20	-21	-22	-24	-21	-21
	Migrant	-	X	X	X	X	-7	-12	-11	-8	-11	-14
	LIEP Waiver	-	X	X	X	X	11	11	13	10	11	12
Race & Ethnicity	Asian nH	-	X	X	X	X	X	20	20	18	20	11
	Asian Hispanic	-	X	X	X	X	X	8	9	8	8	
	Asian Hispanic Disparity	-	X	X	X	X	X	-12	-11	-10	-12	X
	Black / African nH	-	X	X	X	X	X	8	9	10	8	0.1
	Black/African Hispanic	-	X	X	X	X	X	4	4	4	4	
	Black/African Hispanic Disparity	-	X	X	X	X	X	-5	-5	-6	-4	X
	Mixed Multiple Races nH	-	X	X	X	X	X	17	18	16	17	3
	Mixed / Multiple Races Hispanic	-	X	X	X	X	X	3	4	3	3	
	Mixed/Multiple Races Hispanic Disparity	-	X	X	X	X	X	-15	-14	-13	-15	X
	Native American or Alaskan nH	-	X	X	X	X	X	5	5	6	5	0.4
	Native American or Alaskan Hispanic	-	X	X	X	X	X	3	3	3	3	
	Nat American or Alaskan Hispanic Disparity	-	X	X	X	X	X	-2	-2	-3	-2	X
	Pacific Islander or Nat HI nH	-	X	X	X	X	X	2	3	3	2	-3
	Pacific Islander or Nat HI Hispanic	-	X	X	X	X	X	3	3	2	3	
	Pacific Islander or Nat HI Hispanic Disparity	-	X	X	X	X	X	0.4	0	-1	0	X
	White nH	-	X	X	X	X	X	14	13	12	14	2
	White Hispanic	-	X	X	X	X	X	3	4	4	3	
	White Hispanic Disparity	-	X	X	X	X	X	-11	-10	-9	-11	X
	Hispanic (No Race)	-	X	X	X	X	X	3	2	5	3	-7.0*
	Average Hispanic Disparity (All Races)	-	X	X	X	X	X	**-5.9**	-5.6	-5.2	-5.9	
Interactions	Hispanic Newcomer	-	X	X	X	X	X	-11	-11	-10	-11	-9
	Hispanic LTEL	-	X	X	X	X	X	6	6	4	6	7
	Hispanic Female	-	X	X	X	X	X	*0.1*	*0.1*	*-0.1*	*0.1*	*0.1*
	Hispanic IEP	-	X	X	X	X	X	2	2	3	2	2
	Hispanic Migrant	-	X	X	X	X	X	8	7	3	8	10
	Hispanic Waiver	-	X	X	X	X	X	-2	-2	-1	-2	-3
RE Parameters	State	-	FE	X	X	X	X	X	FE	X	X	X
	State | District	-	X	FE	X	X	X	X	X	FE	X	X
	State | District | School	-	X	X	X	X	X	X	X	X	X	X
	State | District | School | Student	-	X	X	X	X	X	X	X	X	X	X
	Residual variance	1242	1221	1161	1240	2025	1060	1034	1023	970	1032	1038
	R-squared (OLS) / ρ (AR1) (xt & mixed)	0.41	0.42	0.42	0.42	0.00	0.49	0.51	0.51	0.53		X
	Constant	277	273	264	275	333	268	260	260	248	260	269

OLS Models

While perhaps overly simplistic with respect to the underlying assumptions in the context of this study, OLS models serve as a good starting point to establish baseline estimates of relationships between focal variables of interest, as the parameter estimates provide a straightforward interpretation with respect to dependencies between predictor and outcome variables. Table 4.1 presents the full list of models estimated under OLS and longitudinal (GLS) specifications.

Naïve Model (0): COVID-19 and Grade fixed-effects

Model 0 establishes a baseline estimate for the impact of COVID-19 on average EL proficiency, while controlling for differences in ELs average' proficiency across grades via inclusion of grade fixed-effects (reported for the final specifications in Appendix A).[72] The parameter estimate implies a difference of almost -7 *CSS* points in post-pandemic years in average EL proficiency. Notably, likely owing to the large samples and consistent differences in average proficiency across grades, the model fit for the *Naïve* model, including only a *COVID binary variable* (and *Grade* fixed-effects) is relatively high, with R-squared estimated at about 0.40. The baseline estimate for the conditional residual variance is 1,242 *CSS* while the model intercept is estimated at about 277 *CSS* points. *Models 0b* and *0c* add *State* and *District* fixed-effects to the *Naïve* model, showing minimal impact on the estimates of the average impact of COVID-19, as well as on other

[72] For academic interest, unconditional, i.e. grade-free estimates are presented in Model 0e. Notably, while this naïve estimate is identical to the reported average difference of -4.9 scale score points (Table 3, Data Section), the R-squared statistic, which is a measure of the fit of the model to the data is practically 0. This also means that the grade fixed-effects are driving a large part of the explanatory power in the regressions, and should be included in all model specifications. Even in the full unconditional model, the parameter estimate for COVID-19 is statistically significant at the 1% level.

general model parameters. *Model 0d* sets a baseline for annual comparisons of aggregate proficiency by substituting the *COVID-19 binary variable* with *Year* fixed-effects. According to this baseline estimate on *Year* fixed-effects, average EL proficiency is 1.6 points lower than in 2022, and still in a gradual post-pandemic decline up to the most recent, 2023 school year. The parameter estimates on the average *impact of COVID-19*, estimated at -4.9 *CSS* (see top row in Table 4.1), is expectedly identical to that reported in Table 3.3.

Model 1. Demographic: no Race and Ethnicity.

Model 1 adds to the *Naïve model* by introducing a basic set of demographic variables as individual-level covariates. These variables are *Time as EL* (measured in Years)*, TELsq* - a quadratic term for *Time as EL*, and binary indicator variables for *Newcomer*, *Long-term EL*, *SLIFE*, *Female*, *IEP*, *Migrant* and *LIEP Waiver*. Model fit statistics, presented at the bottom of Table 4.1, suggest that inclusion of these variables substantially improved the predictive power of the model. R-squared increased to 0.49, while the total residual variance decreased to 1060 *CSS*, compared to 1242 *CSS* of the *Naïve* model. The parameter estimate of the *COVID-19* binary variable decreased only slightly by about one scale score point, showing an average decline of -8 *CSS* in the post-COVID-19 period compared to that of before the pandemic.

The estimates on *TEL = 9* and *TELsq = -0.5* (rounded to -1) from *Model 1* can be interpreted to mean that for each (calendar) year spent as an English Learner students record an average gain of about ten scale score points. Because this estimate is consistent across model specifications and slightly higher at about 10 *CSS* in more precise and final model specifications, this estimate of “about an average annual gain of

10 CSS per year" can be used as a reference to evaluate the magnitude of other estimated coefficients in the model. For example, in the context of *Model 1*, the previously reported average impact of *COVID-19* estimated at about -8 *CSS* points is nearly equal to the jointly estimated effect of *TEL* and its *quadratic term (9 + -0.5)*, setting the preliminary estimate of the average impact of COVID-19 equal to approximately a (calendar) year of instructional time for ELs enrolled in supplementary language support services.

The *Time as EL* effect is estimated separately from the difference of -4 *CSS* estimated for students who could potentially be identified by *Long-term* EL status. *Newcomer* students report average proficiency scores at about 9 *CSS* lower than ELs who have taken the ACCESS assessment in prior years (and thus have been identified as EL in prior years), thus offsetting the equally-estimated learning gains as predicted by the coefficient of the *TEL* variables for the first year in language support programs. The parameter estimate of the coefficient on *SLIFE* is also negative, implying that for each year of additional age difference between the students' and the grade-cohort average age there is an associated -7 *CSS* difference in average proficiency. *Female* students, on average, outperform their male peers by about 5 overall composite scale score points, while *Migrant* students' scores, on average, are about -7 *CSS* lower compared to *non-Migrant* ELs. Students who have a *LIEP Waiver* outperform their peers regularly enrolled and receiving language support programs in schools by 11 *CSS*. The largest absolute difference across a demographic subgroup is estimated for dually-identified students; in this model *ELs with IEPs* report proficiency scores that are 20 *CSS* below their peers without *IEP* identification. Again, all parameter estimates, except for that for *Hispanic*

Female students, are non-trivial, and with a few notable exceptions, are consistent across increasingly more flexible model specifications.

Notably, *Model 1* does not include two focal demographic variables: those identifying students' *Ethnicity* and *Race*, as well as interaction terms of demographic variables with *Hispanic* ethnicity. Discussing these models, the estimated coefficients, and potential impacts on EL proficiency in the context of *Model 1* is informative, however, as (jumping ahead) all of these coefficients are consistent in signs, and most are similar in magnitude across the various model specifications. Moreover, the inter-related changes in model coefficients across various specifications inform the nuanced relationships between individual-level factors and their interactions and how they relate to EL subgroup proficiency. Because the estimated coefficients for the demographic variables are consistent between *Model 1 and Model 2,* I next focus on the additional explanatory information provided by those interaction terms.

Model 2. Intersectional: Ethnicity X Race, Hispanic Interactions

Next, *Model 2* completes the set of included demographic variables by introducing students' *Ethnicity* and *Race*, along with interactions of variables capturing students' identification as *Newcomer*, *LTEL*, *Female*, *IEP*, *Migrant*, and *LIEP Waiver* with *Hispanic Ethnicity*. *Race* and *Ethnicity* are also interacted, creating 14 ethno-racial categories, with a baseline group of students who have no race reported and are identified as *not Hispanic*. While the model fit, as measured by the R-squared value of 0.51, is not much higher compared to the analogous estimate of 0.49 from the previous model, all of the coefficients of the additional covariates are estimated very precisely except for that of the subgroup of *Hispanic* and *Female* ELs, based on the interaction of *Ethnicity* and *Gender* identifiers.

The parameter estimates on variables identifying ELs' ethno-racial categories, based on the interaction of *Ethnicity* and *Race,* vary substantially from 2 to 20 *CSS*, providing preliminary evidence of substantial differences between the outcomes of various ethno-racial subgroups of students. Summarizing these disparities, the last row in the '*Race and Ethnicity*' section in Table 4.1 the provides the *'Average Hispanic Disparity for All Races'* (i.e., controlling for race), by assigning an equal weight to the disparity between *Hispanic* and *non-Hispanic* students estimated for each of the seven reported races. [73]

For example, according to *Model 2* estimates, as indicated in the row labelled *'Asian: Hispanic Disparity'*, *Not Hispanic* and *Asian* students are the subgroup with highest EL proficiency, with scores that are on average 20 *CSS* points higher than the baseline subgroup's average proficiency; moreover, these students outperform their *Hispanic* and *Asian* peers by 12 *CSS* points. *'Average Hispanic Disparity for all Races'*, as estimated by the OLS *Model 2*, is nearing -6 *CSS*.

Given the interaction of several covariates with Hispanic ethnicity some of the model parameters are affected more than others, when compared to those of *Model 1*. The average estimates of *COVID-19, TEL, TELsq, SLIFE, Female, IEP,* and *LIEP Waiver* adjusted only marginally, while those for *Newcomer*, *LTEL*, and *Migrant* ELs change

[73] As explained in the data section, technically there are only 5 reported races. Following the focus of the Intersectionality lens, the sixth is constructed from data for a more nuanced understanding of ethno-racial differences, (e.g. Mixed/Multiple Races), while the seventh, the baseline category is identified by missing data. The Hispanic disparity estimated for the 'No race reported' subgroup of students is that between students who report *Hispanic* and *no Race* and those who report *not Hispanic* and *no Race*.

substantially in magnitude. More specifically, after the introduction of the ethno-racial identification variables and *Hispanic* interactions, *Newcomers'* average proficiency is higher by an estimated 7 *CSS*, while that for *LTELs* is lower by 4 *CSS* (estimated at -2 and -8, respectively). *Migrant* students' average proficiency is also substantially different as compared to the parameter estimate of -7 in *Model 1*, now estimated at 12 *CSS* points lower than that for *non-Migrant ELs*. However, these differences in magnitudes of "main effects" coefficients across *Model 1* and *Model 2* are not surprising, given the relatively large and precise estimates in *Ethnicity* interaction parameters, which signal substantial differences in the context around proficiency estimates for *Newcomers, LTELs, Migrant EL*s of *Hispanic* ethnicity. More specifically, according to the updated and more precise estimates of *Model 2*, while *Newcomer* ELs' proficiency in general is estimated only 2 scale score points lower than those who have been ELs for longer than a year, this difference is much more pronounced for *Hispanic Newcomers*, for whom the difference is estimated at an additional 11 *CSS* lower. This relationship is reversed for *LTELs*; ELs who have been in a language support program for over 5 years report lower scores (8 *CSS* lower compared to *non-LTELs*), but for *Hispanic* students this difference is positive, estimated at 6 *CSS*. The same is true for *Migrant ELs*; the estimated difference of -12 *CSS* between *Migrant* and *non-Migrant* students is smaller for *Hispanic ELs*, estimated at (-12 + 8 =) -4 *CSS*. These inter-related changes in estimated model coefficients highlight important dependencies between covariates, supporting their inclusion in more complex models for a more precise estimation of model parameters and a more nuanced understanding of factors that may be related to ELs proficiency.

Auxiliary Models 2b-e

Four alternate specifications are explored within Model 2, which includes the full set of covariates, but importantly does not account for sources of variation in proficiency that may be due to institutional-level factors. *Models 2b* and *2c* examine EL proficiency in the context of the full set of individual-level covariates while adding *State* and *District* fixed-effects. Notably, the addition of either changes the focal parameter estimates (including that of the impact of COVID-19) only slightly in case of *State* fixed-effects, and marginally more when adding *District* fixed-effects. The biggest changes in parameter estimates from the baselines reported in main *Model 2* are observed for ELs identified as *Migrants* and those with *LIEP Waivers* (3-4 scale score points higher in *Model 2c*), for ELs identified with *IEPs* (3 scale score points lower) and for *Hispanic Migrants* (5 scale score points lower in *Model 3c*). These changes in parameter estimates imply potential differences in ways ELs with *Migrant, IEP,* and *LIEP Waiver* identification, and especially those identified as *Hispanic*, are clustered within, and served by different districts. Importantly, the inclusion of *State* and *District* fixed-effects has a very small effect on the coefficients related to ethno-racial identification, with the *Average Hispanic Disparity* decreasing slightly from 5.9 to 5.6 to 5.2 *CSS*, when *State* and *District* fixed-effects are included, respectively. The model fit also increases very slightly, with the R-squared estimate unchanged after the inclusion of *State* fixed-effects and increased from 0.51 to 0.53 when *District* fixed-effects are included. Total model variance and the constant estimated at 970 and 248 scale score points, respectively, are at their lowest in the OLS specification with *District* fixed-effects. *Model 2c* also establishes a ceiling estimate for the average impact of *COVID-19*, estimated at -8.2 *CSS*.

Next, *Model 2d* examines differences in EL's average proficiency across time using *Year* fixed effects, with the 2017 school year serving as the baseline. The model estimates for annual differences from the 2017 average are 2.2, 0.7, and -0.5 *CSS* for the pre-COVID years of 2018, 2019 and 2020, and -5.8, -7.0, and -8.2 *CSS* for the post-COVID-19 years of 2021, 2022 and 2023, respectively. These estimates of annual learning losses due to COVID-19 are even lower compared to those in *Model 0d*, showing that the inclusion of the demographic variables has surfaced an even larger impact of the pandemic than an unconditional comparison would estimate. Notably, as shown by the model constant, model fit, and residual variance estimates, the parameter estimates of auxiliary *Model 2e* are very similar to that of the main *Model 2,* suggesting that *Year* fixed-effects and the *COVID-19* binary variable are identifying and quantifying similar impacts. This implies that the average proficiency of ELs has been relatively stable (controlling for individual-level factors) when considered within pre-COVID-19 and post-COVID-19 periods separately, and gives additional credence to the estimated coefficient of COVID-19 and other parameters estimates from the main models.

Finally, *Model 2e* examines the effect of decoupling *Ethnicity* from *Race*, by including these in the regressions as separate and independent variables, each with its own reference group (not *Hispanic* for *Ethnicity*, and *No Race Reported* for *Race)*. The parameter estimates of this model that ignores the ethno-racial intersectionality of ELs are close in magnitude to most of the coefficients of the main *Model 2*. Notable differences are observed when comparing the adjustments in the parameter estimates for *Ethnicity*, *Race,* and *Hispanic interaction* variables. Moreover, according to both the total residual variance, estimated at 1038 *CSS*, the model constant, estimated at 269 *CSS*, and R-squared, estimated at 0.50, *Model 2e* is inferior to *Model 2* (1034, 260, and 0.51) in terms of statistical precision and extent of accuracy in variance decomposition. This lends additional credence to the *Intersectional* approach (e.g., interacting *race* and ethnicity*)* in the context of the EL academic outcome data under examination and within the framework of OLS regressions.

While informative and revealing, the parameter estimates from OLS models should be treated with caution, as they may potentially be imprecise without a more flexible model structure that more directly factors in temporal- and institutional-level variations in EL proficiency: many ELs take the ACCESS assessment multiple times across years, and students are clustered within specific schools, districts, and states. To examine these sources of variation in EL proficiency, the next subsections introduce more advanced and flexible model specifications.

Longitudinal Models

Main Model 3

A longitudinal *Model 3* is presented in the first columns of Table 4.2.[74] This model, specified in Equation 2, accounts for the temporal variation in students' scores as many EL students take the ACCESS assessment multiple times in their academic journey towards academic English proficiency.

The average *impact of COVID-19* is estimated at -5.3 *CSS* according to the longitudinal *Model 3*. This estimate is about two scale score points higher (less negative) compared to the corresponding estimate of -7.7 in the OLS counterpart *Model 2*. Other

[74] For reference and convenience of comparison, I also present the main OLS specification of Model 2.

parameter estimates also differ in magnitude, but importantly, none change signs, signaling consistently estimated relationships between these covariates and EL proficiency. Notably, some of the parameter estimates of coefficients that have adjusted more substantially across OLS and Longitudinal model specifications, such as *Time in EL, Newcomer* and *LTEL,* are temporal in nature, and are thus expected to change with a more direct accounting of this source of variation in *Model 3*. For example, as indicated by the substantially smaller magnitude of the parameter estimate for the *Newcomer* variable, estimated at -7 *CSS* as compared to that of -2 *CSS* in *Model 2*, and the substantially larger (less negative) coefficient of *LTEL*, estimated at -4 *CSS* as compared to -8 *CSS* in *Model 2* imply important differences in ways the now-included repeated nature of observations further explains variations in student scores.

Factoring in the temporal variation in students' scores also substantially adjusts the estimated coefficients of some other demographic variables. For example, the average proficiency of *ELs with IEPs* was estimated at -21 *CSS* lower as compared to ELs without *IEPs* in the more naïve OLS *Model 2*; this disparity is still sizeable, but much smaller (less negative) in *Model 3*, estimated at -12 *CSS*. A similar finding is reported for *Migrant ELs*: the coefficient for this variable increased from -12 in *Model 2* to -5 in *Model 3*. The opposite is true for the *LIEP* variable; in the longitudinal *Model 3,* ELs with LIEP waivers outperform their "regularly-enrolled" peers by 5 *CSS*, as compared to the estimate of 11 *CSS* in *Model 2.*

Variables capturing *ethno-racial* identification (i.e., interaction of *Race* and *Ethnicity*) are also differentially affected by the inclusion of the temporal dimension of variation in EL proficiency. The estimated coefficients for some subgroups increased, while others decreased as compared to the OLS *Model 2*. The average *Hispanic Disparity* is slightly higher in *Model 3*: -7 *CSS* in *Model 3* vs -6 *CSS* in *Model 2*. *Hispanic* interaction variables with *Newcomer*, *LTEL*, and *Migrant* also adjusted, as all of the parameters have much smaller estimated magnitudes in *Model 3* compared to *Model 2*.

Table 4.2: Parameter estimates of longitudinal and mixed-effects main and auxiliary models.

	REGRESSION FAMILY	OLS	Longitudinal				MIXED (Random-effects/Hierarchical)										
	Models Numbers *(DV = CSS)*	M2	M3	*M3b*	*M3d*	*M3e*	M4	M5	M6	M7	*M7b*	*M7d*	*M7e*	*M7f*	*M7g*	*M7h*	*M7i*
	Model Name, Levels and Structure; Type of residuals and variance	+ Full Demo: ExR, Interactions	RE GLS, AR1	+ State FE	+ Year FE ('17 baseline)	Demo: E + R	2-Level RE, AR1	3-Level RE, AR1	4-Level RE, AR1	5-Level RE, AR1	4L: + State FE	4-Level: + Year FE	5L, R+E	5L, Naïve	5L, Demo	5L, no '21	5L, no AR1
	Average Impact of COVID-19	-7.7	-5.3	-5.3	2.9; 2.8; 3.3 -2.4; -2.3; -1.7	-5.4	-5.5	-6.6	-6.6	-6.6	-6.6	2.5, 1.8, 1.7 -4.8, -5.1, -4.6	-6.7	-4.7	-6.7	-7.0	-6.5
Demographics	Time as EL (in Years)	9	11	11	11	11	11	10	10	**10**	10	10	10	X	10	10	10
	TEL Squared	-1	-1	-1	-1	-1	-1	-1	-1	**-1**	-1	-1	-1	X	-1	-1	-1
	Newcomer	-2	-7	-7	-7	-8	-8	-7	-7	**-7**	-7	-7	-7	X	-9	-7	-7
	Long-term EL	-8	-4	-4	-4	-4	-4	-4	-4	**-4**	-4	-4	-4	X	-3	-5	-5
	SLIFE (Δ from cohort avg age)	-6	-5	-5	-5	-5	-5	-4	-4	**-4**	-4	-4	-4	X	-4	-4	-4
	Female	5	7	7	7	7	7	6	6	**6**	6	6	6	X	6	6	6
	IEP	-21	-12	-12	-12	-12	-11	-16	-16	**-16**	-16	-16	-16	X	-15	-17	-16
	Migrant	-12	-5	-4	-5	-5	-4	-4	-4	**-4**	-4	-4	-4	X	-3	-4	-3
	LIEP Waiver	11	5	6	5	6	5	5	5	**5**	5	5	5	X	4	5	5
Race & Ethnicity	Asian nH	20	20	20	20	16	20	15	15	**15**	15	15	11	X	X	15	15
	Asian Hispanic	8	6	6	6		6	6	7	**7**	7	7		X	X	7	7
	Asian Hispanic Disparity	-12	-14	-14	-14	X	-14	-8	-8	**-8**	-8	-8	X	X	X	-8	-8
	Black / African nH	8	7	7	7	4	7	8	8	**8**	8	8	5	X	X	9	8
	Black/African Hispanic	4	1	1	1		1	2	3	**3**	3	3		X	X	3	3
	Black/African Hispanic Disparity	-5	-6	-6	-6	X	-6	-6	-6	**-6**	-6	-6	X	X	X	-6	-6
	Mixed Multiple Races nH	17	14	14	14	4	12	11	11	**11**	11	11	3	X	X	12	11
	Mixed / Multiple Races Hispanic	3	*0*	*0*	*0*		*-1*	*1*	*1*	***1***	*1*	*1*		X	X	*2*	*1*
	Mixed/Multiple Races Hispanic Disparity	-15	-14	-14	-14	X	-13	-10	-10	**-10**	-10	-10	X	X	X	-10	-10
	Native American or Alaskan nH	5	1	1	1	-0.3	*0.0*	2	2	**2**	2	3	0.4	X	X	3	3
	Native American or Alaskan Hispanic	3	-1	-2	-1		-2	1	1	***1***	*1*	*1*		X	X	*2*	*1*
	Nat American or Alaskan Hispanic Disparity	-2	-2	-3	-2	X	-2	-1	-1	***-1***	-1	-2	X	X	X	-1	-2
	Pacific Islander or Nat HI nH	2	1	1	1	-1	1	2	2	**2**	2	2	-0.5	X	X	3	2
	Pacific Islander or Nat HI Hispanic	3	*-1*	*-1*	*0*		*-1*	*0.3*	*0.4*	***0.4***	*0.4*	*1*		X	X	*1*	*0*
	Pacific Islander or Nat HI Hispanic Disparity	0.4	-2	-2	-2	X	-2	-2	-1	**-1**	-1	-1	X	X	X	-1	-2
	White nH	14	*11*	*10*	*11*	3	*10*	8	8	**8**	8	8	2	X	X	9	8
	White Hispanic	3	*0*	*0*	*0*		*0*	2	2	**2**	2	2		X	X	3	2
	White Hispanic Disparity	-11	-10	-10	-10	X	-10	-6	-6	**-6**	-6	-6	X	X	X	-6	-6
	Hispanic (No Race)	3	*0*	-1	0	-5.9*	-1	2	2	**2**	2	2	-2.7*	X	X	3	2
	Average Hispanic Disparity (All Races)	**-5.9**	**-7.0**	-7.0	-6.9		**-6.8**	**-4.4**	**-4.3**	**-4.3**	-4.3	-4.3		X	X	-4.1	-4.4
Interactions	Hispanic Newcomer	-11	-4	-4	-4	-3	-3	-4	-4	**-4**	-4	-4	-3	X	X	-5	-4
	Hispanic LTEL	6	1	1	1	1	1	2	2	**2**	2	2	2	X	X	2	2
	Hispanic Female	*0.1*	*-0.1*	*-0.1*	*-0.1*	*-0.1*	*0.0*	*-0.2*	*-0.3*	*-0.3*	*-0.3*	*-0.3*	*-0.3*	X	X	*-0.3*	*0.0*
	Hispanic IEP	2	2	2	2	1	2	2	2	**2**	2	2	1	X	X	2	2
	Hispanic Migrant	8	3	3	3	4	3	1	*1*	***1***	*1*	*1*	1	X	X	*1*	*1*
	Hispanic Waiver	-2	-2	-2	-2	-3	-2	-1	-1	**-1**	-1	-1	-2	X	X	-1	-1
RE Parameters	State	X	X	FE	X	X	X	X	-	**29**	FE	29	33	26	45	27	28
	State | District	X	X	X	X	FE	X	X	100	**82**	81	82	83	81	100	78	82
	State | District | School	X	X	X	X	X	X	200	113	**113**	113	111	115	224	133	111	117
	State | District | School | Student	X	696	683	692	701	808	654	653	**653**	653	653	655	815	672	750	767
	Residual variance	1034	353	355	355	353	452	375	375	**375**	375	376	376	428	375	267	263
	R-squared (OLS) / ρ (AR1) (xt & mixed)	0.51	0.29	0.29	.7;.4;.5; 0.29	0.28	0.47	0.38	0.38	**0.38**	0.38	0.38	0.38	0.43	0.38	0.38	X
	Constant	260	257	251	255	262	257	258	261	**260**	249	259	264	265	265	259	260

Again, these differences in estimated parameter estimates are not overly surprising, as subgroups of EL students representing different and intersecting individual identities display disparate average outcomes. For example, dually-identified students scores are typically much lower compared to their peers without *IEP* identification. Given this, it is easy to see that OLS models are essentially under- or over- estimating disparities by pooling all data together, and not taking into account the fact that many observations of proficiency are recorded for the same student, albeit differently over time.

Fit statistics for *Model 3* suggest a similar, if not higher explanatory power as reported by the estimates of overall R-squared of 0.50, with the corresponding within- and between- estimates of R-squared, reported for longitudinal models, estimated at 0.70 and 0.42, respectively. In *Model 3* the total conditional variance, estimated at 1050 *CSS*, is comprised of the estimated student-level variation (696 *CSS*) and conditional residual variance (353 *CSS*). While just a little higher than that in *Model 2* (1034 *CSS*), the explained (versus random) part of the variation in *Model 3* is much higher, while the model constant is just a little smaller, estimated at 257 *CSS*, compared to that of 260 *CSS* in *Model 2*. The autocorrelation coefficient ρ, capturing the magnitude of serial correlation is estimated at 0.29. These statistics all suggest that the parameters estimated under the longitudinal framework are more robust and realistic compared to those estimated under the simpler OLS framework.

Auxiliary Models 3b-e

Despite the better (than OLS) fit, *Model 3* still fails to account for institutional levels of variation, i.e., *School*, *District*, and *State* fixed-effects. While *Model 3b*, similar to its OLS counterpart *Model 2b*, factors in state-level variation by including *State* fixed-effects, a respective model *'3c'* is not feasible in the longitudinal specification due to computational power limitations. Meanwhile, the inclusion of *State* fixed-effects has a negligible impact on parameter estimates compared to those reported by the main longitudinal *Model 3*.

Model 3d estimates the longitudinal model using year dummy variables, which is equivalent to including *Year* fixed-effects as in *Model 2d*. The coefficients on the year dummy variables, estimated at 2.9, 2.8, and 3.3 *CSS* for the pre-COVID-19 school years 2018, 2019, and 2020, and at -2.4, -2.3, and -1.7 *CSS* for the post-COVID-19 years 2021, 2022, and 2023, are in reference to the school year 2017. These estimates are similar to the difference in the *impact of COVID-19* binary estimate and report a smaller estimated impact of the pandemic on average EL proficiency than that estimated in the OLS counterpart model. Promisingly, in contrast to the OLS counterpart *Model 2d*, the estimate of the 2023 school year is larger than that of 2022, implying that under the more flexible and more precise longitudinal / GLS model specification the average English proficiency of ELs is showing an upwards trend.

Finally, *Model 3e* decouples *Race* and *Ethnicity*, providing independently estimated relationships and coefficients for students' racial and ethnic identifiers. All the other covariates in the model are minimally different from those reported in the main *Model 3*. However, there are noteworthy differences between the estimates of secondary *Model 3e* and baseline *Model 3*. For example, the parameter estimates for *Asian* and *Black/African ELs* are respectively 4-5 *CSS* points higher in *Model 3e* (compared to the reference group) estimated at 16 and 4 *CSS*, as compared to that of 11 and 0.1 in *Model 2e*. *Pacific Islander* and *White* ELs' reported proficiency is also slightly higher in the longitudinal model specification. Ignoring the ethno-racial intersectionality in Model *3e*, *Hispanic* ELs are reporting average proficiency levels 6 *CSS* lower compared to non-*Hispanic* ELs, which mimics the slight increase in most of the parameters identifying racial subgroups.

Summary of Longitudinal Models

In sum, the estimates from longitudinal models corroborate the evidence from OLS models and build on them by introducing more flexible and realistic model assumptions. The focal variables of interest change slightly across specifications, with the *COVID-19 impact* estimate slightly smaller, and the *Average Hispanic Disparity* estimate slightly higher. The time-related variables of *TEL, TEL_sq; LTEL* and *Newcomer* see the largest adjustments across specifications and are more precisely estimated in the longitudinal/GLS specification (*Model 3*). Importantly, while it constitutes an improvement over OLS *Models 0-2* the longitudinal specification does not account for the nesting of students in specific schools, districts, and states. Findings from Mixed-effects models, presented in the next section, address this shortcoming via inclusion of random-effects at the respective levels.

Mixed Models

Findings from main mixed-effects *Models 4*-7, and the secondary *Models 7b-7i* are presented on the right-hand side part of Table 6.3 under a unifying header shaded in green. Main models *M4* through *M7* are also highlighted in different shades of green, signaling the increasing levels of nesting in more complicated and flexible models. Random-effects parameters, populated for appropriate levels of included nesting for various model specifications are presented in the bottom part of Table 6.3, and can be compared to the variance estimates obtained from both OLS (Table 4.2) and longitudinal models.

Model 4. Two-level: Students ↓Time

Model 4 offers a mixed-effects specification that directly accounts for the repeated nature of *Student-level* observations across time. As such, the conceptual design, levels, and structures of included variation are the same across the longitudinal/GLS *Model 3* and the mixed-effects *Model 4*. While the estimation methods differ across the longitudinal and mixed-effects specifications, these similarities of the two models are apparent in comparing the fixed and random parameter estimates from *Model 3* to those of *Model 4*. What is different across the two specifications is how the two variance components are decomposed. More specifically, the longitudinal/GLS model appears to be more precise in this task, as both the components of *Student-level* (across-time) variance and residual (random, or unexplained) variance estimates are smaller under this specification. The total error variance is also smaller in the longitudinal model, estimated at 1050 *CSS*, compared to that of the mixed-effects model 4 estimated at about 200 *CSS* higher, at 1260 *CSS*. This is likely due to the difference in the estimated autocorrelation coefficient, which is likely overestimated at $\rho = 0.47$ in the mixed-effects *Model 4.*

Model 5. Three-level: Schools ↓ Students ↓Time

I build on *Model 4* by adding *School* random-effects in addition to the *Student* random-effects. Notably, *Model 5* is the first specification enabling examination of school-level variations in EL proficiency, and how its inclusion in model adjusts estimated regression parameters. Comparing the estimating parameters of fixed and random coefficients across *Model 3* and *Model 4* reveals some interesting patterns. While most parameter estimates, such as that on the average *impact of COVID-19*, *Time as EL*, *Newcomer*, *SLIFE*, *Female*, and *Hispanic interactions* are only slightly different (about one scale score point lower or higher) several others have changed substantially with the inclusion of school-level variation in the estimation. For example, while the *IEP* coefficient ranges from about -11 to -12 *CSS* in *Model 3* and *Model 4,* it is lower by 5 *CSS*, estimated at -16 *CSS* in *Model 5*.

Some of the estimates on ethno-racial disparities have adjusted similarly: *Asian Not Hispanic* students report average proficiency that is 15 *CSS* higher (versus the baseline group of *no Race, not Hispanic*) as compared to the 20 *CSS* difference reported in *Model 4* (thereby decreasing the reported *Asian Hispanic Disparity from -14 to -8 CSS*). Also noteworthy is the smaller *Hispanic Disparity* for *While ELs, estimated at* -6 in *Model 5* compared to -8 in *Model 4*. These two changes in parameters drive the smaller estimated *Average Hispanic Disparity* at -4.4 vs -6.8 in *Model 4.*

Adding *School-level* variation into the model also changes the random-effects parameters in important ways. First, despite the more complicated nesting structure, the total variance as decomposed in *Model 5* and estimated at 1229 *CSS* is smaller compared to that in *Model 4.* The residual variance is also quite smaller (by 80 *CSS*), estimated at 375 *CSS*. Second, the *Student-level* variance estimated at 808 *CSS* in *Model 4* has been further decomposed into 654 *CSS* capturing *Student-level* variance, and 200 *CSS* capturing *School-level* variance. Finally, the autocorrelation coefficient, estimated at $\rho = 0.38$ in *Model 5* is also smaller than its counterpart in *Model 4* implying that some of the “inertia” effects (of student’s scores correlation across time) are absorbed and better predicted by the *School*- random-effects. These changes in random effects parameters signal a better fit in *Model 5* and support the inclusion of school-level random effects. Further, the non-trivial changes in regression coefficients after the inclusion of *School* random-effects signal potentially important differences in ways different subgroups of ELs are being served across schools (discussed in more detail in Chapter 5).

Model 6. Four-level: Districts ↓ Schools ↓ Students ↓Time

Model 6 introduces *District* random-effects into the estimation, elevating the number of nested levels to four. While impressive from a modeling perspective, the change in model coefficients is smaller than the integer-level rounding of *CSS* points can capture.[75] In other words, introducing *District-level* variation to (on top of) the mixed-effects regression estimated in *Model 5* does not result in an adjustment for the previously estimated relationships, when “only” *School-level* and *Student-level* variation were explicitly included. Interestingly (and perhaps jumping ahead), this same relative invariance to added institutional levels of variation is observed when *State* random-effects are added in *Model 7*, as described in the next section.

Random-effects estimates shown at the bottom of Table 4.2 present the only small change across the parameter estimates that can be observed from adding *District* random-effects in *Model 6*. More specifically, the *School-level* variation in *CSS*, estimated at 200 points in *Model 5* is decomposed into 113 and 100 points, representing the respective *School*- and *District*-level variance estimates in *Model 6*. While all other random-effects parameters have not changed, the total conditional error variance, estimated at 1142 *CSS* in *Model 6* (100+113+653+375) is substantially smaller than that

[75] I could not find any research in education that applies empirical models with more than three levels of nesting. Moreover, there are no studies that investigate the multi-level nesting of English Learner students.

in *Model 5* estimated at 1229 *CSS*. This implies that despite the apparent stability of the fixed parameter estimates, the “random part” of the mixed-effects model estimates benefited from the inclusion of *District* random-effects. In other words, the inclusion of district-level nesting improved the overall model fit and precision of variance decomposition, but not necessarily its fixed parameter estimates.

Model 7. Five-level: States ↓ Districts ↓ Schools ↓ Students ↓Time

Model 7 is the final specification that models the relationship between EL proficiency and individual-, temporal-, school-, district-, and state-level factors, as shown in Equation 3 (Chapter 3). Importantly, similar to the case of *Model 6*, the fixed parameter estimates of this model are again not substantially different after including the highest level of nesting, i.e., *State* random-effects. The identical (after rounding) parameter estimates for all fixed coefficients across *Model 5* to *Model 6* to *Model 7* imply that the estimated relationships and dependencies have stabilized with respect to additional levels of nesting hierarchies, and signal model saturation with respect to the hierarchical nesting.

Similar to the transition from *Model 5* to *Model 6*, the random effects parameter estimates of the *Model 7* changed only with respect to the further decomposition of the highest included level of variation. *District-level* variance, estimated at 100 *CSS* in *Model 6* is further partitioned into 82 and 29 *CSS* points, representing the respective *District-* and *State-level* variance estimates in *Model 7.* Signaling saturation of the hierarchical nesting from the standpoint of variance decomposition, the total conditional variance in *Model 7*, estimated at 1153 *CSS* is slightly higher than its counterpart in the four-level *Model 6*, estimated at 1142 *CSS*.

Since the fixed (non-random) parameter estimates are identical across *Models 5-7*, I will not discuss these separately for *Model 7*. Instead, I outline the findings on parameter estimates from secondary specifications *M7b-M7i,* and compare the latter to both the Main specification in *Model 7,* and to their respective OLS and GLS counterparts.

Auxiliary Models 7b-i

Model 7b replaces the *State* random-effect with a *State* fixed-effect. *Model 7b* is therefore a four-level mixed-effects model with *State* fixed-effects added to adjust for the state-level variance. Thus, the random-effects estimates of variance parameters of *Model 7b* can be compared to those of *Model 7* and *Model 6.* Expectedly, the four-level fixed-effects specification yields in the smallest reported model constant, reported at 249 (because this approach effectively "differences out" *CSS* points from the fixed part of the model). *Model 7b* is equivalent to *Model 7* for practical purposes but is much faster to execute (5 days of computer runtime compared to three weeks). This is important for future research, as it supports the use of four-level models with more flexible variance decomposition, especially relevant when more data is added to the models with the completion of the 2023-24 ACCESS administration.

Model 7d implements yet another four-level specification, now modeling the lowest level (i.e., temporal variation) of hierarchical variance through *Year* fixed-effects. Other than the differently-decomposed *COVID-19* impact, the parameter estimates from *Model 7d* are otherwise almost identical to *Model 7*. Notably, similar to the findings from the Longitudinal/GLS counterpart, findings from *Model 7d* on annual differences in average proficiency (in reference to school year 2017) also show an upward trend for the most recent school year of 2023. While the latter is perhaps the most positive finding this study can offer, we can be cautiously optimistic that the trend may be reversing.

Model 7e follows its OLS and longitudinal (GLS) counterparts and decouples *race* and *ethnicity* in a five-level random-effects specification. Akin the OLS and longitudinal specifications, comparisons with the main *Model 7* reveal minimal differences in parameter estimates, apart from those capturing ethno-racial identification. Notably, the difference between *Hispanic* and *non-Hispanic ELs'* proficiency as quantified by the "decoupled" *Model 7e* is estimated at -2.7 CSS, while the main *Model 7* suggests a more accurately (both theoretically and empirically) estimated, and larger "*Average Hispanic Disparity*" of -4.3 CSS. Therefore, examining EL proficiency under an *Intersectional* lens surfaced additional, larger inequities that would otherwise remain invisible.

To compare with the baseline estimates of *Model 0*, *Model 7f* provides a *Naïve* estimate of the *impact of COVID-19* by removing all fixed parameters from the model while leaving the four-level nesting intact. The coefficient is smaller in the OLS *Naïve* model, implying that the 2.2 *CSS* difference between the estimated parameters is likely absorbed by the institutional-level effects. Also absorbed by these random effects is the impact of now-omitted fixed effects of the variables capturing ethno-racial identification of ELs. While the *State*- and *District*-level variance components have not changed across specifications, both the student- and school-levels of variation are substantially larger. It is difficult to ascertain the reasons behind these dynamic and interrelated changes. The differential clustering of specific ethno-racial subgroups of students across schools, the different ways in how various schools serve these students, and how the performance of these students has been affected across time could be some of the factors that could be driving these changes in the estimates.

Model 7g removes the *ethno-racial* variables from the main model, leaving only the basic demographic variables. As can be seen by comparing the estimates of *Model 7g* to the main *Model 7* parameters, removing the ethno-racial variables does not affect the fixed parameter estimates substantially. On the other hand, similar to the *Naïve* model random-effects estimates, the random-effects parameter estimates in *M7g* are much larger than those in the main *Model 7.* Different from the *Naïve* model, however, is how the institutional-level variability adjusts from including only the basic demographic variables. More specifically, the changes in variance parameter estimates are more proportional, with the *State*- and *District-level* variance increasing along with more modest (than in the Naïve model) increases in *Student*- and *School-level* variance. This implies that the inclusion of *ethno-racial* and *Hispanic* interaction variables helps explain variations in EL proficiency along all levels. In other words, it is corroborating evidence that there are non-trivial differences in how EL students representing different ethno-racial backgrounds and multiple intersectionalities are served by different schools, districts, and states. This opens doors for future research to examine these differences by adding random slopes for focal groups of research interest, to enhance and improve the higher-level variance decomposition.

Model 7h replicates the main specification in *Model 7,* by removing the school year 2021 from the analysis due to concerns that its potentially selected nature, along with the smaller sample size and higher measurement error in this “during COVID-19” academic year, could distort some of the estimates. However, alleviating these concerns exclusively from a robustness standpoint, the average *COVID-19 impact* estimate is (slightly) even more negative, estimated at -7.0 *CSS* as compared to the more precise and stable estimate of about -6.6 *CSS* in *Models 5-7.* While the estimates of the fixed parameters do not change substantially, this is not true of the variance estimates specified in the random effects. More specifically, the residual variance in *Model 7h* is 100 *CSS* lower, estimated at 267 *CSS,* while the *Student-level* variance is higher by about the same amount, estimated at 750 *CSS*, compared to those in the main *Model 7*. A similar adjustment or coefficients can be observed when removing AR1 error structure. In other words, either removing school year 2021 in *Model 7g* or misspecifying the error variance structure by not accounting for the moderately high serial correlation in *Model 7i* inflates the within-student variation on the account of the residual variance. This lends additional support to including the 2021 data and autoregressive errors in the final mixed-effects models.

Finally, *Model 7i,* while estimated under a naive error variance structure, enables the calculation of an important statistic – the residual Intra-cluster correlation coefficient (ICC), which measures the degree of clustering of observations (of EL proficiency) within groups (i.e., states, districts, schools, and students). It also represents the degree of variability in ELs' scores between groups. The *ICC* for the state-, district-, school-, and student-levels is (very precisely) estimated at 0.02, 0.09, 0.18, and 0.79, respectively. These estimates suggest that, after adjusting for district- and school-effects that likely absorb some of the higher (state)-level variation, there is moderate variation in EL proficiency across districts and schools, but not so much across states. The estimate of 0.79 for the student-level variation is high, but likely an overestimate, as it fails to account for the fact that many of the student-level observations are correlated across time.

Summary of Main OLS, Longitudinal, and Mixed-effects Model Results

Table 4.3 collects the main regression models and parameter estimates together, removing the secondary models for a side-by-side comparison of focal variables.

Table 4.3: Parameter estimates of main regression models.

	REGRESSION FAMILY	ORDINARY LEAST SQUARES (OLS)			LONGITUDINAL	MIXED (Random-effects/Hierarchical)			
	Models Numbers *(DV = OCSS)*	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Model Name, Levels and Structure; Type of residuals and variance	Naïve (Grade FE)	+ Demo: no R&E	+ Full Demo: RxE, Interactions	Random Effects GLS, AR1	+ Student: 2-Level RE, AR1	+ School: 3-Level RE, AR1	+ District: 4-Level RE, AR1	+ State: 5-Level RE, AR1
	Average Impact of COVID-19	-6.9	-8.0	-7.7	-5.3	-5.5	-6.6	-6.6	-6.6
Demographics	Time as EL (in Years)	X	9	9	11	11	10	10	**10**
Demographics	TEL Squared	X	-1	-1	-1	-1	-1	-1	**-1**
Demographics	Newcomer	X	-9	-2	-7	-8	-7	-7	**-7**
Demographics	Long-term EL	X	-4	-8	-4	-4	-4	-4	**-4**
Demographics	SLIFE (Δ from cohort avg age)	X	-7	-6	-5	-5	-4	-4	**-4**
Demographics	Female	X	5	5	7	7	6	6	**6**
Demographics	IEP	X	-20	-21	-12	-11	-16	-16	**-16**
Demographics	Migrant	X	-7	-12	-5	-4	-4	-4	**-4**
Demographics	LIEP Waiver	X	11	11	5	5	5	5	**5**
Race and Ethnicity	Asian nH	X	X	20	20	20	15	15	**15**
Race and Ethnicity	Asian Hispanic	X	X	8	6	6	6	7	**7**
Race and Ethnicity	**Asian: Hispanic Disparity**	X	X	-12	-14	-14	-8	-8	**-8**
Race and Ethnicity	Black / African nH	X	X	8	7	7	8	8	**8**
Race and Ethnicity	Black/African Hispanic	X	X	4	1	1	2	3	**3**
Race and Ethnicity	**Black/African: Hispanic Disparity**	X	X	-5	-6	-6	-6	-6	**-6**
Race and Ethnicity	Mixed Multiple Races nH	X	X	17	14	12	11	11	**11**
Race and Ethnicity	Mixed / Multiple Races Hispanic	X	X	3	*0*	*-1*	*1*	*1*	***1***
Race and Ethnicity	**Mixed / Multiple Races: Hispanic Disparity**	X	X	-15	-14	-13	-10	-10	**-10**
Race and Ethnicity	Native American or Alaskan nH	X	X	5	1	*0*	2	2	**2**
Race and Ethnicity	Native American or Alaskan Hispanic	X	X	3	-1	-2	1	1	***1***
Race and Ethnicity	**Nat AM or Alaskan: Hispanic Disparity**	X	X	-2	-2	-2	-1	-1	***-1***
Race and Ethnicity	Pacific Islander or Nat HI nH	X	X	2	1	1	2	2	**2**
Race and Ethnicity	Pacific Islander or Nat HI Hispanic	X	X	3	*-1*	*-1*	*0*	*0*	***0***
Race and Ethnicity	**Pac Isl or Nat HI: Hispanic Disparity**	X	X	0.4	-2	-2	-2	-1	**-1**
Race and Ethnicity	White nH	X	X	14	*11*	*10*	8	8	**8**
Race and Ethnicity	White Hispanic	X	X	3	*0*	*0*	2	2	**2**
Race and Ethnicity	**White Hispanic: Disparity**	X	X	-11	-10	-10	-6	-6	**-6**
Race and Ethnicity	**Hispanic Disparity (No Race)**	X	X	3	*0*	-1	2	2	**2**
Race and Ethnicity	**Average Hispanic Disparity (All Races)**	**X**	X	**-5.9**	**-7.0**	**-6.8**	**-4.4**	**-4.3**	**-4.3**
Interactions	Hispanic Newcomer	X	X	-11	-4	-3	-4	-4	**-4**
Interactions	Hispanic LTEL	X	X	6	1	1	2	2	**2**
Interactions	*Hispanic Female*	X	X	*0.1*	*-0.1*	*0.0*	*-0.2*	*-0.3*	*-0.3*
Interactions	Hispanic IEP	X	X	2	2	2	2	2	**2**
Interactions	Hispanic Migrant	X	X	8	3	3	1	*1*	***1***
Interactions	Hispanic Waiver	X	X	-2	-2	-2	-1	-1	**-1**
RE Parameters	State	X	X	X	X	X	X	X	**29**
RE Parameters	State | District	X	X	X	X	X	X	100	**82**
RE Parameters	State | District | School	X	X	X	X	X	200	113	**113**
RE Parameters	State | District | School | Student	X	X	X	696	808	654	653	**653**
RE Parameters	Residual variance	1242	1060	1034	353	452	375	375	**375**
	R-squared (OLS) / *ρ (AR1) (xt & mixed)*	0.41	0.49	0.51	.7,.4,.5; *0.29*	*0.47*	*0.38*	*0.38*	***0.38***
	Constant	277	268	260	257	257	258	261	**260**

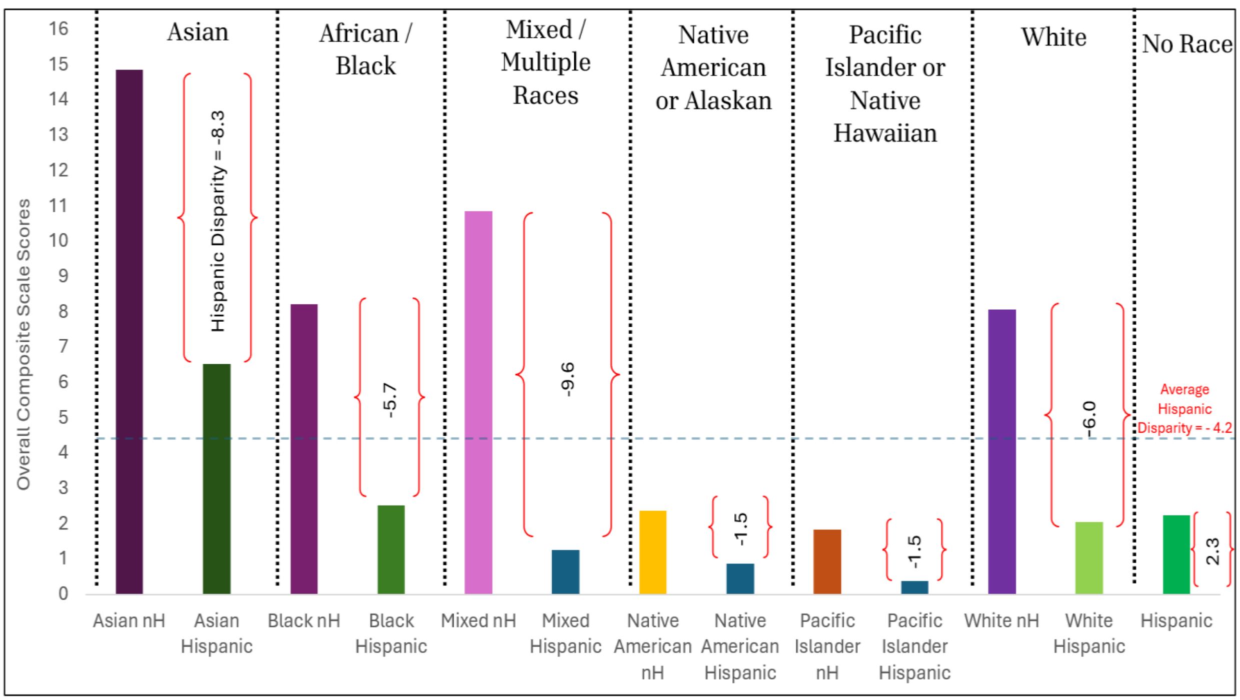
The parameters on the focal variables of interest estimated across various model specifications present a consistent picture of a large, sustained, and differential impact of the pandemic on various EL subgroups. Due to the common scale across the various specifications and the same underlying analytic sample, it is useful to compare coefficients across models for a more nuanced understanding of the relationships and dependencies across and between variables and levels. The dynamic and interrelated changes in parameter estimates across various specifications illustrate how various modeling assumptions affect these relationships. The population-level samples underlying the analysis coupled with the high precision of the estimates reported in Tables 4.1, 4.2 and 4.3 provide a comprehensive map of statistical relationships and can serve as a high-level blueprint in future analyses.

Findings from regression models show that the average impact of the COVID-19 pandemic on EL proficiency is estimated at -6.6 *CSS* points. While within the context of the overall theoretical range of *CSS* points (100-600) for an individual student this estimate does not seem large, its magnitude is more telling when compared to the other parameter estimates of the model. For example, as compared to the coefficient of *TEL* estimated at about 10 *CSS*, the average impact of *COVID-19* can be restated as equal to about 0.66 (calendar) years of *Time as EL*. In this light, some of the disparities uncovered by the parameter estimates of models are quite unsettling. For example, the disparity between ELs with and without *IEP* identification is estimated at 16 *CSS*, which would be roughly equal to the impact of two and half COVID-19 pandemics! In *TEL* terms, this disparity implies that in terms of English language proficiency, *dually-identified* ELs are on average about a year and half behind their peers without *IEP* identification. [76]

[76] Another way to evaluate the reported differences and impacts is through the standard deviation units, using the data provided in Table 3.2. However, this is also somewhat imprecise, as these deviations are estimated for an entire

Similar concerns are uncovered with respect to disparities between ethno-racial subgroups. Figure 4.1 shows that controlling for the multitude of multileveled factors included in *Model 7*, average proficiency of select *ethno-racial* subgroups, such as *Asian*, and especially *Not Hispanic Asian* ELs, or *Not Hispanic African/Black* and *Not Hispanic White ELs* is much higher than that of *both Hispanic* and *not Hispanic Native American/ Alaskan EL*s, *Pacific Islander / Native Hawaiian ELs.* Disparities by *Hispanic* status for each of the *Races,* and for '*All Races on Average*" are also substantial and show that *Hispanic* students' scores are 4.2 *CSS* lower than those of their *non-Hispanic* peers.

Figure 4.1: Ethno-racial disparities in average EL proficiency



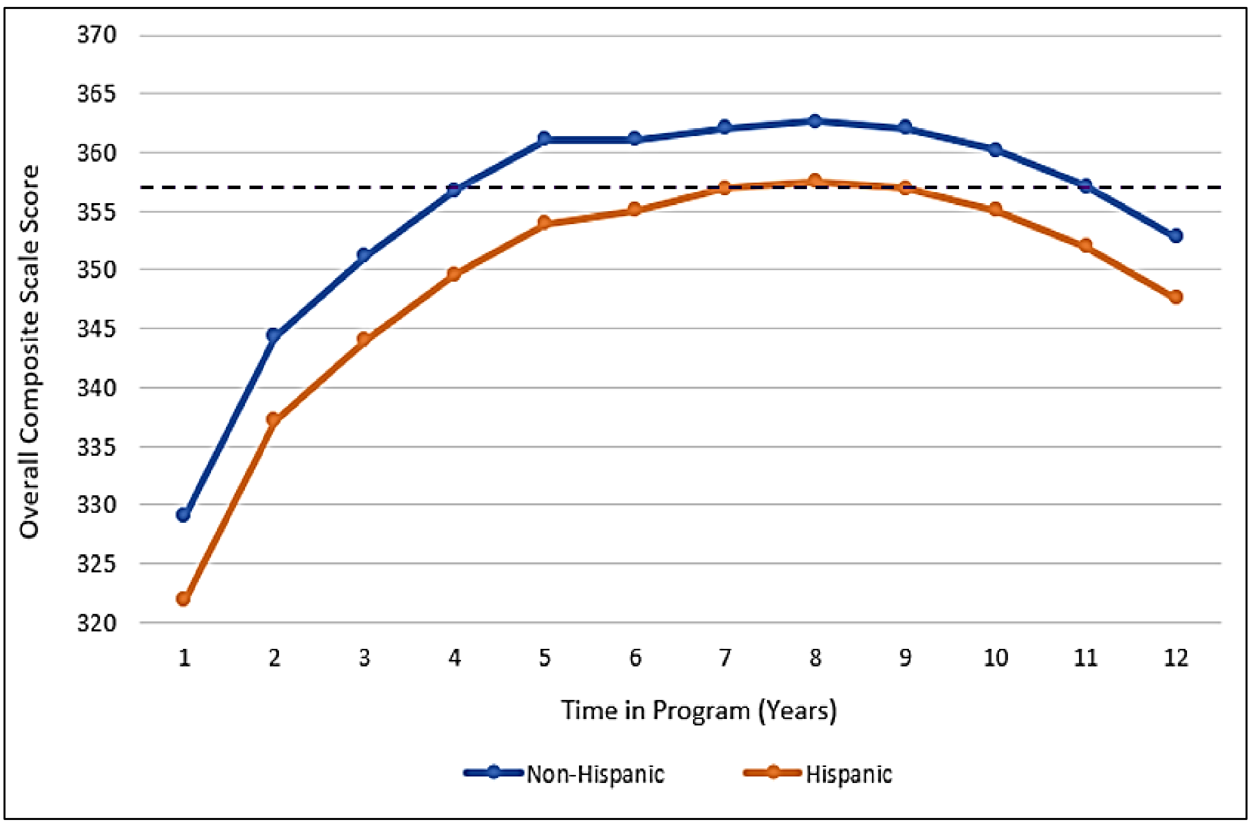
To further highlight the impact of these disparities by *Hispanic* identification, Figure 4.2 presents EL language development trajectories, based on the predicted parameter

grade-level, and also vary across grades. Using the pooled (across-grades) estimate of standard deviation of about 45 CSS, a difference of 10 CSS would imply an effect size of about 0.25 standard deviations.

estimates of *Model 7* on *Time as EL* (and its quadratic). Average temporal effects of *Newcomer* and *LTEL* and respective *Hispanic Interactions*, as well as the average *impact of COVID-19* are included in this comparison, and trajectories are estimated for ELs enrolled in an 'average grade' (using the estimates of *Grade* fixed-effects).

Disparities in average proficiency across ethnic identification become more apparent when comparing the predicted average English language development trajectories of *Hispanic* versus *not Hispanic English Learners* over time, as given in Figure 4.2.

Figure 4.2 Predicted disparities in language development by Hispanic identification.



For example, Figure 4.2 shows that after three years from initial identification, *Hispanic ELs'* average proficiency lags about a year behind that of *non-Hispanic ELs'*.

Four years after initial identification, close to the peak of the subgroup proficiency across years (as indicated by the dashed line), *Hispanic* ELs are about three years behind their *non-Hispanic* peers.

Finally, using the random effects parameters and total residual variance estimates from the final specification in Model 7, it is possible to arrive at an overall decomposition of variance in ELs' average proficiency, as shown in Figure 4.3.

Figure 4.3 Sources of variations in EL proficiency.

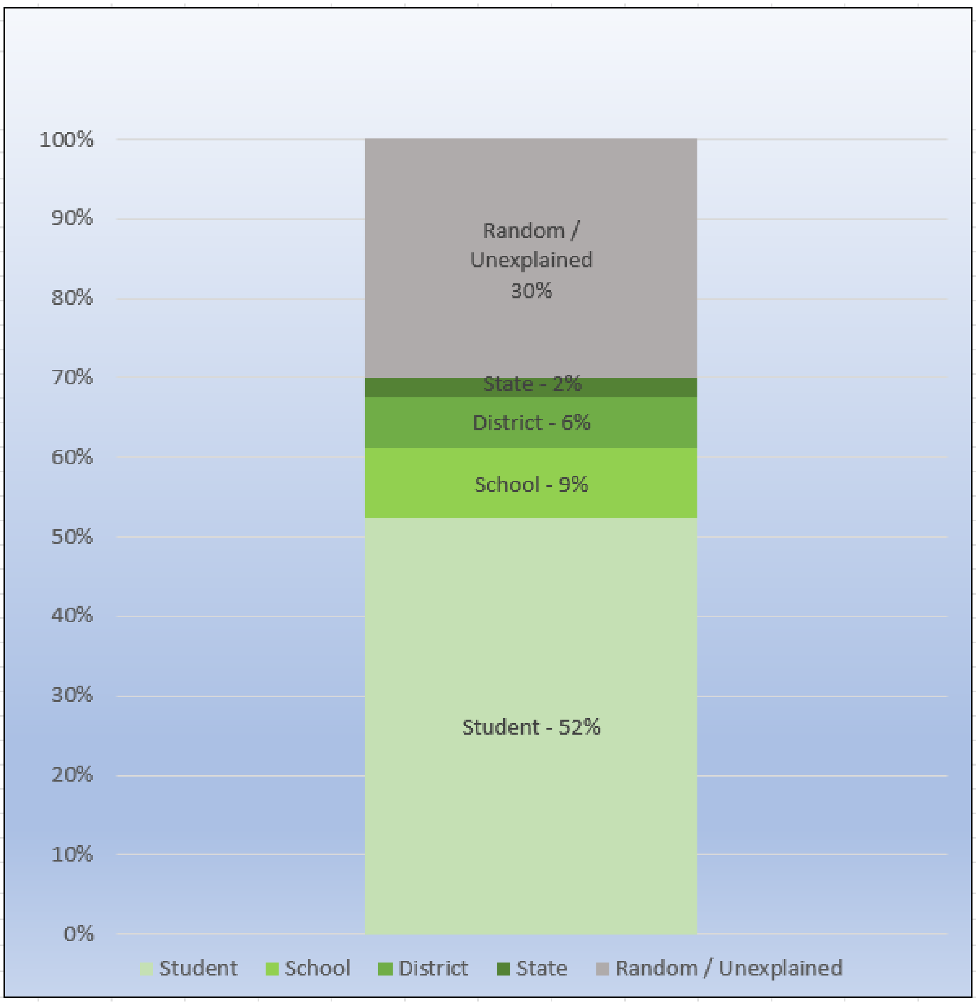


Figure 4.3 shows that according to Model 7 random-effects parameters, close to 70% of the variation in EL scores is sourced to State (2%), District (6%), School (9%), and Student (52%) - level factors, while 30% of the variance remains unexplained.

The estimates from the various regression models and the final specification *Model 7* address the first research question, by confirming a large and sustained impact of COVID-19 on EL proficiency, and documenting sizeable disparities across ethno-racial, and other subgroups. The precise and consistent estimated relationships between individual-level factors and EL proficiency in the context of temporal and institutional-level variables provide a detailed roadmap on how EL proficiency was impacted and shaped by these multilevel factors. However, these estimates are limited in the sense that they quantify average relationships for the entire timespan under consideration (2017-2023). While illuminating with respect to estimates of the average overall impact the pandemic and overall differences and disparities in subgroup outcomes, these estimates don't offer much insight with respect to how these relationships were affected more recently, after the pandemic caused the disruptions in the education of ELs. To explore how the pandemic impacted these relationships, I generate and compare estimates for pre- and post-COVID-19 periods *separately*, replicating the specification presented in *Model 7*. These results are provided in the next section.

Impact of the COVID-19 Pandemic on EL Disparities

In this section, I replicate the specification presented in *Model 7* to compare the differences in average EL proficiency by subgroup of interest for the pre- and post-COVID-19 periods. Thus, the relationships between covariates and EL proficiency are estimated separately for 2017–2020 and 2021–2023 and are presented in Table 4.4. Because the difference in parameter estimates across the pre- and post-COVID-19 specifications are relatively small (but still of research interest), I report decimal points for all of the estimated fixed coefficients. Reporting how these relationships have been impacted by the pandemic, the final column of Table 4.4 presents the changes in post- to pre-COVID-19 estimates for all the variables included in the models, including the parameters of variance estimated for student, school, district, and state-levels. In the rest of this section, I explore each demographic variable of interest in turn. I conclude the section with a discussion of the patterns that emerge across focal EL subgroups.

Time as EL

The parameter estimate on *Time as EL* approximates, in *CSS* units, the average amount of academic English acquisition for an EL student, controlling for the multitude of factors and covariates included in the analysis. According to the estimated difference of 1.2 *CSS*, this amount slightly increased for ELs in the post-COVID-19 era, while the quadratic term, capturing diminishing returns, is slightly more negative. Due to the smaller timespans (four and three school years instead of seven) under consideration resulting from partitioning of the analytic sample into pre- and post-COVID eras, it is difficult to ascertain the source of these small differences in the impact of *TEL*.

Table 4.4: The impact of COVID-19 on predictors of EL proficiency.

	Dependent Variables / Period	Pre-COVID	Post-COVID	Post-Pre
Demographics	Time as EL (in Years)	9.8	10.9	1.2
	TEL Squared	-0.6	-0.7	0.0
	Newcomer	-7.9	-2.8	5.1
	Long-term EL	-4.5	-5.4	-0.9
	SLIFE (deviation from cohort avg age)	-4.4	-4.1	0.3
	Female	6.2	5.6	-0.5
	IEP	-18.0	-17.6	0.4
	Migrant	-2.8	-4.8	-2.0
	LIEP Waiver	5.9	5.0	-0.8
Race and Ethnicity	Asian nH	14.5	16.7	2.2
	Asian Hispanic	6.9	7.2	0.2
	Asian: Hispanic Disparity	-7.6	-9.6	-2.0
	Black / African nH	7.2	10.9	3.7
	Black/African Hispanic	3.1	3.3	0.2
	Black/African: Hispanic Disparity	-4.1	-7.6	-3.5
	Mixed Multiple Races nH	11.7	12.6	0.9
	Mixed / Multiple Races Hispanic	2.3	1.9	-0.4
	Mixed / Multiple Races: Hispanic Disparity	-9.4	-10.7	-1.3
	Native American or Alaskan nH	2.5	4.0	1.5
	Native American or Alaskan Hispanic	1.5	0.9	-1
	Nat AM or Alaskan: Hispanic Disparity	-0.9	-3.1	-2.1
	Pacific Islander or Nat HI	2.3	2.8	0.6
	Pacific Islander or Nat HI Hispanic	*1.4*	*0.9*	*-0.5*
	Pac Isl or Nat HI: Hispanic Disparity	-0.9	-1.9	-1.0
	White nH	8.3	10.4	2.1
	White Hispanic	3.1	2.0	-1.1
	White Hispanic: Disparity	-5.2	-8.4	-3.2
	Hispanic Disparity (No Race)	3.0	2.4	-0.6
	Average Hispanic Disparity (All Races)	**-3.6**	**-5.5**	**-2.0**
Interactions	Hispanic Newcomer	-4.6	-4.9	-0.3
	Hispanic LTEL	2.1	4.5	2.4
	Hispanic Female	*-0.1*	*-0.3*	*-0.2*
	Hispanic IEP	1.8	2.5	0.7
	Hispanic Migrant	1.0	*0.8*	*-0.2*
	Hispanic Waiver	*-1.5*	*-0.1*	*1.4*
RE Parameters	State	41	23	-18
	State|District	80	84	4
	State|District|School	104	94	-10
	State|District|School|Student	582	641	59
	Residual variance	392	353	-39
	ρ (AR1)	0.39	0.34	-0.05
	Constant	263	251	-12

Newcomer ELs

The average disparity by *Newcomer* status, es estimated for the entire timespan including pre- and post-COVID-19 periods and reported in Table 4.3, was calculated at about 7 *CSS*. There are notable changes to parameter estimates for this variable when the relationship is estimated for pre- and post-COVID-19 periods. More specifically, the difference of the parameter estimate is comparatively the largest among focal subgroups (see the last column in Table 4.4). The overall proficiency of *Newcomer ELs* (taking the ACCESS test for the first time), while still slightly lower compared *non-Newcomer ELs* (-2.2), is 5 *CSS* higher after the pandemic. Importantly, however, this improvement in *Newcomer* proficiency, does not manifest the same way for *Hispanic* students. As indicated by the slightly more negative coefficient for the *Hispanic and Newcomer* variable, the average language proficiency of *Hispanic Newcomers* does not show a similar post-COVID-19 improvement.

Long-term ELs

The average disparity by *LTEL* status, as estimated for the entire timespan including pre- and post-COVID-19 periods and reported in Table 4.3, was calculated at about 5 *CSS*. Unlike the findings for *Newcomer ELs,* however, *Long-term ELs'* average proficiency has declined by about 1 *CSS*, from an estimated -4.5 (pre-COVID-19) to -5.4 *CSS* (post-COVID-19). Interestingly, the change in the *Hispanic LTEL* interaction variable from post- to pre-COVID-19 periods is also sizeable, indicating that these students, as opposed to *not Hispanic LTELs*, are recording slightly higher scores after the pandemic.

SLIFE

According to the small parameter estimate of 0.3 *CSS* for this variable reported in the last column of Table 4.4, there is only a slight difference between the parameter estimates for *SLIFE* pre and post pandemic. In other words, interruptions to students' formal education have a similar estimated negative impact on average EL proficiency in both pre- and post-COVID-19 periods.

Gender

While *Female* EL students outperform their male peers by about 6 *CSS* points on average, this estimate is slightly smaller in the post-COVID-19 years. The *Hispanic* interactions between *Gender* and *Hispanic* identification, similar to the estimates in the overall specification for the entire timespan, are still statistically not significant, implying that *Hispanic Female* ELs do not report either higher or lower scores on average, as compared to *Hispanic ELs*, or *Female EL* students.

IEP Status

The average proficiency outcomes for English learners with *IEPs* are substantially lower compared to their peers without disabilities. Similar to the post- to pre- COVID-19 changes in parameter estimates for *SLIFE* and *Female*, the changes in the *IEP* parameter estimate are small (estimated at 0.5 *CSS*). *Dually-identified Hispanic* students' average scores are slightly higher post-COVID-19, estimated at 2.5 *CSS* as compared to 1.8 *CSS* in pre-COVID-19 years.

Migrant Status

Migrant ELs' average proficiency has decreased after the pandemic by 2 *CSS* from an estimated -2.8 *CSS* to -4.8 *CSS*. *Hispanic Migrant ELs* do not report statistically different average proficiency levels before and after the pandemic.

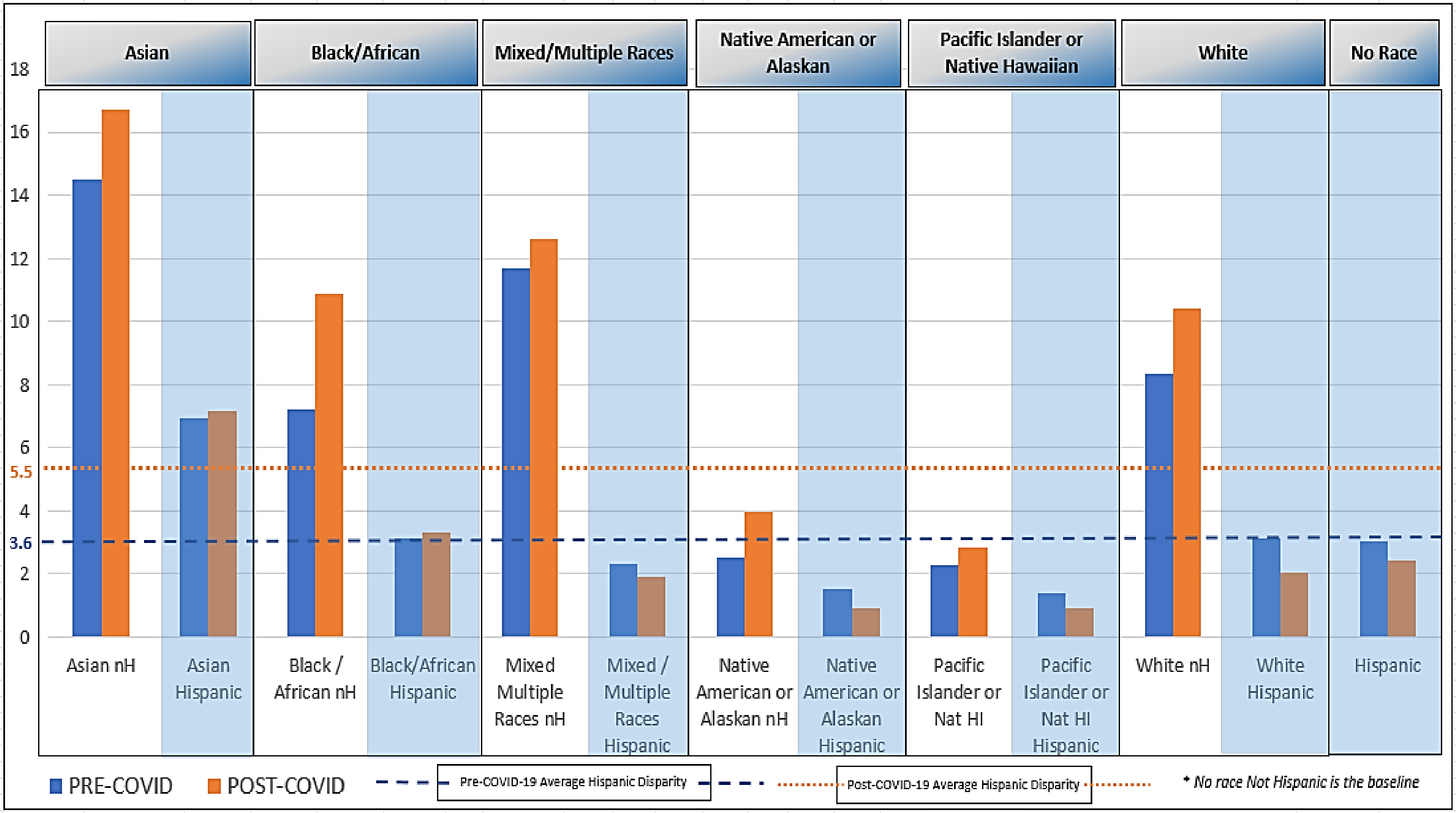
LIEP Waiver

EL students with *LIEP Waivers* generally outperform their peers who are enrolled in supplementary language support programs at schools. This difference is slightly smaller in post-COVID-19 years, but still sizeable and estimated at 5 *CSS*. The respective term with the *Hispanic* interaction is not statistically significant in the post-COVID era.

Race and Ethnicity

Examining differences between ethno-racial subgroups' average performance in pre- and post-COVID-19 years provides some notable findings. While there are some improvements in the average proficiency of several subgroups in post-COVID-19 years (as compared to the *no Race, not Hispanic* subgroup), the sizeable, negative and higher parameter estimates for *Hispanic Disparities* derived for each of the *Ethnicity* and *Race intersections* as well as for *All Races*, show that disparities in average proficiency outcomes by *Hispanic* ethnicity increased after the pandemic. Notably, higher average proficiency in the post-COVID-19 years is reported for all *non-Hispanic* ethno-racial subgroups, with *Not Hispanic Asian* (2.2 CSS), *Not Hispanic African/Black* (3.7 *CSS*) and *Not Hispanic White* (2.1 *CSS*) ELs reporting relatively larger increases in average *CSS* in the post-COVID-19 years. Figure 4.4 provides a visual of the changes to ethno-racial disparities by subgroup due to the COVID-19 pandemic, as summarized in Table 4.4.

Figure 4.4. Ethno-racial disparities in average EL proficiency before and after COVID-19.



Student, School, District, and State-level Random Effects

Presented in the last column at the bottom part of Table 4.4 are notable differences in the estimated variance parameters before and after the COVID-19 pandemic. Variance parameters due to *State*- and *School*-level factors are lower in the post-COVID-19 period. This is especially true for the estimate of the *State* random effect, which is almost halved in magnitude. *District*-level variance, on the other hand, slightly increased from 80 *CSS* to 84 *CSS*. The residual variance is slightly higher in the pre-COVID era compared to its estimate in the post-COVID-19 years of 2021-2023, implying that variations in EL proficiency are less random after the pandemic.

Summary of Findings on the average impact of COVID-19 on EL Disparities

In sum, the findings from replicating *Model 7* with pre- and post-COVID-19 data indicate that the pandemic impacted relationships between predictor and dependent variables (including disparities by subgroup and institutional-level impacts) in varied, but consistent and somewhat predictable ways. Many of the subgroup disparities increased, while others decreased. Importantly, the most consistent decreases in average language proficiency were estimated for *Hispanic* English learners, who represent the majority of students both in the analytic sample and nationally.

CHAPTER FIVE: DISCUSSION

Introduction

Akin to the uncovered disparities that were precisely quantified by the advanced and multilevel regression models and presented in the previous chapter, the theory and development of statistical methods enabling these models also have a deep-rooted history in systemic, institutional, and individual racism (Russell, 2023). Francis Galton, Karl Pearson, and Ronald Fisher – each credited with the advancement of statistical theory and models that underly the exponentially more advanced and flexible successors that are implemented in this study though cutting-edge statistical software and computer processing power – all held racist and eugenicist beliefs that permeated the implications they drew from their analyses (Tabron & Thomas, 2023; Russell, 2023). The methods and tools they created, while (arguably) objective in their nature, were developed for the purpose of attempting to validate unfounded and self-serving white supremacist beliefs, and to move forward racist arguments about heredity and selective breeding (e.g., Clayton, 2021). Following the latter:

"That skull measurements could indicate differences between races – and by extension, differences in intelligence or character – was almost axiomatic to eugenicist thinking. Establishing those differences in a way that appeared scientific was a powerful step towards arguing for racial superiority" (p. 144, in Russell, 2023).

The theory and implementation of statistical methods and regression models have come a long way since then. Studies are more careful in disentangling issues of correlation from causation, as problematic issues prevalent in observational data such as sample selection, omitted variables, and simultaneity are more commonly and explicitly addressed or at least acknowledged in educational research and measurement. Econometricians use terms like *endogeneity* to remind us that correlation is not causation, and that the underlying mechanisms of relationships and dependencies quantified and expressed by estimated regression coefficients need theoretical grounding, further exploration, and perhaps most importantly, careful interpretation. However, while many things have changed from the time of these misguided scientists, racist narratives and repressive ideas still permeate the discourses surrounding immigrants or “foreigners,” as well as persons racialized Black, Hispanic, or Asian, among others. At times, these narratives are even amplified at very high political levels. Therefore, it is important to reiterate once again that the disparities uncovered by the regression analyses for many of the intersectional student groups <u>*are not signals of causal impacts, or effects*</u> of racial, ethnic, or other identity, or their intersections. Simply assigning a person (student) to specific subgroups, especially when the assignment itself is based on inequitable rules defined by the (educational) system, does not *cause* a change in their score: rather, it is the lived (academic) experiences that systematically differ among racialized and ethicized student groups that contribute to disparities (Russell, 2023). These identity markers serve as mere proxy variables that are related to these systematically different experiences and are then captured and quantified by the coefficients of regression models.

Echoing these points, Roberts (2011) draws on genetic analyses from the Human Genome Project that show a greater variation among people with recent African heritage and among people with recent European heritage than there is variation between these groups, leading to the obvious conclusion that there is nothing genetically inherent in people that supports grouping them in racial categories based on biology, and that "race *itself* is an invented political grouping. Race is not a biological category that is politically charged. It is a political category that has been disguised as a biological one" (Roberts, 2011, p.4). Another example illustrating this point is the famous quote credited to the editors of the journal Nature Biotechnology (2004), that "[s]cientifically, race is a meaningless marker of anything. Pooling people in race silos is akin to zoologists grouping raccoons, tigers, and okapis on the basis that they are all stripey." (p. 903) [77]

Further, astrophysicist Neil De Grasse Tyson opines[78] that applying the scientific concept of *albedo* – instead of the discrete categories of race as applied by individuals, institutions, and systems – would perhaps be less harmful in the context of race-based wars, genocides, ethnic "cleansings" and conflicts that have plagued human history since similar-looking groups of people have been able to congregate and militarize. Resonating with the goals of the *anti-categorical* approach of *Intersectional* complexity analysis (McCall, 2005), Dr. Tyson's hope is that this might (still?) promote the understanding that this shared common characteristic (of reflectivity) exists in a full and continuous spectrum, rather than dividing us into discrete and "colorful" categories, which are applied to label, divide, and classify humans into "us" and "them."

Meanwhile, the findings from this study presented in the last chapter provide consistent evidence of a large, persistent, yet differential impact of the COVID-19

[77] Editorial, Illuminating BiDil, 23 Nature Biotechnology 903, 903 (2005)
[78] Video link: "Neil deGrasse Tyson Explains Albedo." https://www.youtube.com/watch?v=pJ0GQYiBg_U&t=141s

pandemic on English Learner students' outcomes, as estimated by multi-level regression models that account for the potential impact of institutional-level factors, i.e., EL students' nesting across WIDA states, districts, and schools. Importantly, these disparities in outcomes are delineated by many of the above-discussed *discretized* racial and ethnic identity categories and their intersections. Further, the pandemic has impacted these disparities in differential, and – for the majority of ELs, i.e., Hispanic language learners – in detrimental ways. The tension between the socially constructed nature of these variables and their "estimation" in the empirical models – despite the attempts to alleviate it through the applications of interaction terms, "fuzzy set logic" and multilevel models, as suggested by *Intersectional* researchers – cannot be easily resolved. However, while the analysis unavoidably relies on ethno-racial categories by including them as "predictor" variables, the overarching purpose of doing so is to highlight disparities at the intersection of racial, ethnic, and other categories, thereby illuminating shortcomings of the educational system, as well as pointing to areas needing remedies within it.

The discussion in this chapter relies on the adopted *Intersectional* lens for a more nuanced understanding of the uncovered disparities in the context of the institutional factors that helped shape them. The next sections present the results around these disparities and provide a discussion guided by the overarching purposes and guiding principles of a critically quantitative analysis.

Results

RQ1: Impact of the COVID-19 Pandemic on Average Proficiency

Addressing research question 1, Tables 4.1-4.3 document precise and consistent estimates of regression coefficients on individual-level variables capturing students' reported demographic, ethno-racial, and other identities, as well as their intersections. These estimates outline important disparities in some EL students' outcomes, as their proficiency continues to remain substantially behind that of their peers, who move more seamlessly towards higher academic English proficiency that leads towards exit from EL status. Furthermore, while there is some evidence that ELs' average scores are trending slightly upwards in the most recent academic school year examined (*Model 7d*), the evidence also highlights that post-COVID recovery has been insufficient and unequally distributed.

Intersectional Overview of Individual-level Differences and Disparities

Examining average proficiency by students' ethnic and racial identification reveals important and substantial differences. For example, findings show that ELs reporting *Asian*, *Black/African, Multiple/Mixed Races* and *White* racial identities, on average, report markedly higher English proficiency scores as compared to *Native American or Alaskan,* and *Pacific Islander or Native Hawaiian* ELs. Meanwhile, the interaction of *Ethnicity* with *Race* provides an *Intersectional* view and highlights important nuances with respect to differences in average English proficiency outcomes across ethno-racial subgroups of EL students. More specifically, these differences in average proficiency are especially salient for ELs who are also identified as *Hispanic* compared to with those who are not, as substantial and varied disparities between these students are reported for each of the subgroups identified by a different *Race*. [79] Recalling the baseline estimate of “about 10 *CSS* per year as EL,” disparities by *Hispanic* ethnicity are also sizeable for *Asian* (-8.3 *CSS*), *Black/African* (-5.7 *CSS*), and *White* ELs (-6.0 *CSS*).[80] On average, “controlling for Race,” *Hispanic EL* students’ scores are 4.3 *CSS* lower compared to that of ELs without *Hispanic* identification (when averaged across the seven reported races). Importantly, this estimate is larger (more negative) than its counterpart of -2.7 *CSS*, reported in analyses that decouple *Race* and *Ethnicity* and consider them separately (*Model 7e*). This implies that examining disparities in outcomes though an *Intersectional* lens has revealed important, additional, and larger disparities for many English Learner students that would otherwise remain invisible and neglected.

Further, examining the interplay between the parameter estimates of “main” and *Hispanic* interaction variables across *Model 1* and *Model 2*, as well as in the final specification in *Model 7* reveals important differences in how specific demographic factors “explain” differences in average EL proficiency for *Hispanic* versus *non-Hispanic* identified ELs. For example, the final specification *Model 7* (Table 4.3) shows that the interactions of Ethnicity with *Newcomer* status reveal non-trivial differences in the average proficiency levels of newly identified, beginner-level ELs that identify as *Hispanic* versus *non-Hispanic*, as the latter subgroup’s reported average proficiency was 4 *CSS* higher.

[79] Curiously, the largest disparity for Hispanic students is reported for ELs identified by Multiple/Mixed races, warranting further inquiries into the demographic and educational characteristics of these language learners.
[80] It is important to recall that these are average “effects”. There are substantial differences, for example by grade-level, as estimated by the precisely estimated grade fixed-effects (Appendix A).

Conversely, *LTEL* students who identified as Hispanic report slightly higher average proficiency as compared to *non-Hispanic LTELs.* Coupled with the above-discussed result of a lower initial proficiency estimated for *Hispanic Newcomers*, this implies differential (average) language development trajectories for *Hispanic* versus *non-Hispanic* students.

A comparison of such differential trajectories predicted by the parameters of Model 7 is given in Figure 4.2. It illustrates that *Hispanic* students start their academic journeys as ELs at lower English proficiency levels, and, on average, never catch up with their *non-Hispanic* counterparts. Further, these disparities in outcomes become larger with time. More specifically, after three years from initial identification, *Hispanic ELs'* average proficiency is about a year behind that of *non-Hispanic ELs*. Worryingly, four years after initial identification, *Hispanic* ELs are about three years behind their *non-Hispanic* peers. Even more worryingly, four years after identification *Hispanic EL*s' average proficiency is close to its peak, shown by the dashed line. Concerningly, these estimates suggest that absent a substantial positive change and systematic improvements in these students' education and academic experiences, many *Hispanic ELs'* proficiency will not reach the level of their *non-Hispanic EL* peers even with additional time in schools as ELs.

Intersectional Overview of the Institutional-level Factors

The tenets of *Intersectionality* also call for considering the simultaneous and potentially differential impacts of dynamic and institutional contexts, by examining student outcomes in a multilevel regression framework (Rusell, 2024). The five-level specification described in Equation 3 and implemented in *Model 7* includes both temporal and institutional levels in its flexibility, and is a first attempt at applying this approach for examining aspects and nuances in EL education (at this scale and scope). The evidence gathered from various specifications provides ample evidence that institutional context matters, notably in different ways, in forming and shaping these students' educational outcomes, and subsequent status as English Learner.

Speaking to this point is the substantial adjustment of the parameter estimates for some of the ethno-racial variables following the inclusion of institutional, i.e., *School*- and *District*-level random effects across models with added hierarchical levels / random effects. For example, the smaller estimate (-4.4 *CSS*) for *Average Hispanic Disparity* (*for All Races*) in *Model 5 (3L: School* **|** *Student* **|** *Time)* compared to that of -6.8 *CSS* in *Model 4 (2L: Student* **|***Time)* may be indicative of the different ways in how schools are set up to serve various ethno-racial subgroups of students. While this finding could be a signal that the educational system may have a (small) overall "equalizing effect" on average disparities in outcomes for the EL population, without additional, more rigorous, and targeted analyses it is difficult to causally attribute the relative over- or under-performance of any of the ethno-racial groups, or disparities in thereof, to either level of the institutional-level effects. This is not only due to the omitted variables shown in the grey zone of the *Theoretical Framework of EL Intersectionality* in Figure 3.1 that could further "explain" EL performance and thus further adjust these coefficients (although there is "only" 30% variance left to "explain;" see Figure 4.3). Perhaps more importantly, the institutional effects themselves are intertwined, and difficult to disentangle. In other words, while it is the inclusion of the *School*-level random effects that impacts the adjustment of the coefficients in *Model* 5, it is important to note that these *School* effects are not completely independent of higher-level *District*- and *State*-effects; the underlying differences (rules, policies, demographics, etc.) driving some of the *State* and *District*-level variation in student scores could already be partially absorbed by the *School*-level effect.[81] Similarly, the omission of these institutional-level random effects (as in *Model 4*) does not imply that that the estimated individual-level coefficients and disparities based on them are completely free of these effects; on the other hand, these institutional effects are likely partly absorbed in the temporal and individual-level effects (with the remainder captured by the model residuals).

Following this logic, the fact that the “fixed” parameters of the model do not change after the additional inclusion of the *District* and *State*-level random effects in *Models* 6 and *7* should not be taken as causal evidence that what states and districts are doing “does not matter” (with respect to how EL proficiency is manifested, as predicted by these covariates). Rather, that the most salient changes to ELs’ proficiency occur at the school-level could be indicative of schools’ close following and implementation of EL policies, procedures and rules, cascading from the federal and state to district and school level. It may also be indicative that other sources of heterogeneity, such as the demographic composition of the states’ and districts’ EL population, is closely reflected in the heterogeneity at the school-level. In other words, this lack of variation at higher levels could mean that, after controlling for the demographics of the EL population (minus SES,

[81] Future research, bolstered by additional computing power and 2024 ACCESS Online assessment data will consider disentangling these higher-level effects through the stepwise exclusion of lower-level variance parameters – a strategy that is currently not feasible due to technical limitations.

among other unobserved factors), schools closely integrate the features and policies stipulated by districts and states that nest them.

Also impacted by the inclusion of institutional-level variables are the coefficients of some of the other demographic variables. For example, the substantial adjustment of -5 *CSS* in the coefficient of the *IEP* variable (identifying EL students with disabilities) after the inclusion of *School* random-effects indicates that there are a select number of schools (and potentially even districts that nest these unique schools) where dually-identified students are receiving appropriate supports enabling high achievement; conversely however, this also implies that there are many more educational settings (schools and districts) where the disparities in educational outcomes between students identified with and without *IEPs* is much more pronounced. A similar finding is uncovered when considering the adjustment of the coefficient on the interaction variable identifying *non-Hispanic* and *Asian, and non-Hispanic and White* EL students, indicating the potential presence of a few, but very effective (with respect to ELs' average proficiency) educational institutions that are well-equipped to serve students at these specific ethno-racial and other intersections. While not causal, these, and other parameter estimates from *Models 4-7* suggest that the inclusion of institutional-level variables informed the estimated relationships in important ways.

RQ2: Impact of COVID-19 on Individual- and Institutional-level effects

Having estimated the average impact of the pandemic on EL proficiency controlling for individual- and institutional-level factors, I now turn to discussing the estimates of the impact of the pandemic on these individual- and institutional-level factors.

Table 4.4 presents the pre- and post-COVID-19 parameter estimates on all covariates included in the analysis. One of the more notable changes with respect to the magnitude of the impact of COVID-19 on EL subgroups is the average reported performance of *Newcomer* students. More specifically, findings show that after the pandemic (i.e., in school years 2021-2023) *Newcomer* ELs tested at substantially higher levels compared to before the pandemic (2017-2020). An important exception to this, once again, are ELs identified as *Hispanic*, for whom average proficiency results in the first year of taking ACCESS Online remained level at about 5 *CSS* lower compared to *non-Hispanic Newcomers*. Conversely, *Hispanic LTELs* reported slightly higher (2.4 CSS) average proficiency levels in the post-COVID-19 era, signaling some improvement in these students' otherwise plateauing proficiency and sparking hope that there may be some upwards trends in the overall proficiency of *Hispanic* students to look forward to in the 2023-24 ACCESS Online data.

The patterns surfacing from examining the demographic variables on EL ethno-racial identification, however, are less promising for *Hispanic* ELs. Namely, while the parameter estimates visualized in Figure 4.4 suggest upwards trends for many of the intersectional subgroups, most if not all of these improvements in average scores are small, and they are reported primarily by ELs of various races that also identify as *non-Hispanic*. Figure 4.4 shows that the most salient recovery after COVID-19 "learning loss" is reported by *Black*, *Asian* and *White* ELs who are not *Hispanic* – students who were already represented in the ethno-racial subgroups of ELs with relatively higher average proficiency. Some positive post-pandemic trends in average proficiency were also observed for *Native American or Alaskan and Pacific Islander or Native Hawaiian ELs* who identified as *Not Hispanic*. Conversely, while the average proficiency of *Asian Hispanic* and *Black/African Hispanic* ELs remained at the same level, *White Hispanic ELs*, *Native American or Alaskan Hispanic ELs, Hispanic* ELs identified as having *Mixed/Multiple Races*, *Hispanic Pacific Islander* or *Native Hawaiian* ELs, and *Hispanic ELs* who didn't report a racial identifier all recorded even lower average proficiency after the pandemic. Reflecting on the disproportional impacts by students' ethnicity, the estimated overall disparity between the average proficiency of *Hispanic* versus *non-Hispanic* identified ELs increased from 3.6 *CSS* before the pandemic to 5.5 *CSS* after the pandemic. While not large in absolute *CSS* terms, this estimate implies a post-COVID-19 increase of 66% in the estimated overall disparity by *Hispanic* identification.

Finally, examining pandemic-induced differences in institutional-level parameters, school and especially state-level variation in EL proficiency was substantially lower in the post-COVID-19 period, while *District*-level variance remained largely unchanged. The school-level variance, on the other hand, was higher in the post-COVID-19 period, signaling that the pandemic has may have increased differences and disparities in students' average proficiency. Coupled with the results on smaller residual error variance implying that differences in EL proficiency are "less random" and more predictable after the pandemic, this suggests that COVID-19 pandemic, and the varied ways that the different levels of the educational system responded to its challenges, have resulted in substantial shifts in the educational experiences and outcomes of many ELs. Unfortunately, as reflected in the consistently negative signs and substantial magnitudes of the parameter estimates of several focal covariates, these shifts are mostly indicating further marginalization for many of WIDA's English Learners.

Implications

The source of these disparities, and the differential ways they have been impacted by the pandemic is a combination of complex factors the examination of which warrants a separate (perhaps mixed-methods) inquiry. Among these factors are substantial shortages and disparities in educational funding. For example, as reported by Darling-Hammond (2007), the wealthiest US public schools spend at least 10 times more than the poorest schools, and these differences contribute to a wider achievement disparity than in virtually any other industrialized country.

The situation has not improved since then, as at the onset of the COVID-19 pandemic the Century Foundation reported that the nation is underfunding education by $150 billion per year compared to what would be necessary to make sure all children, and especially those from ethnic minority and low-income backgrounds, have access to quality education.[82] The report (2020) further highlighted that: (a) districts with high concentrations of students living in poverty were more likely to have funding disparities, and these students experienced significantly larger funding disparities than wealthier districts; (b) districts with high concentrations of Hispanic and Black students had larger funding disparities and were more likely to have funding gaps to begin with than majority white districts; (c) districts with the largest funding disparities had a high concentration of

[82] The Century Foundation: Closing Americas Education Funding (2020). Retrieved from: https://tcf.org/content/report/closing-americas-education-funding/

Hispanic students, and (d) large variations and disparities also exist at the state-level, including underfunded districts even in high-funding states. The report predicted that as the pandemic constrained state and district budgets even more, additional cuts to public education may have exacerbated these gaps, concluding:

> *"Inequity in public education is not a natural occurrence, but rather the result of funding choices. Decades of disinvestment in public education at the state and federal level have a cost, and it has primarily come at the expense of Latinx, Black, and low-income students. As protests across the country call into question how our policies affect communities of color and where we choose to direct our resources, policymakers have the power to make different choices that advance equity, rather than exacerbate inequality."*

While these figures and disparities refer to funding of students' education overall, they are certainly much more magnified and pronounced for English Learners' education, which the literature and many reports describe as severely underfunded (Villegas, 2023; Frengi, 2021).[83,84]

Supporting these predictions, this study uncovers and documents large, persistent, and growing disparities within many English leaner subgroups' proficiency outcomes, while accounting for potential differences in how states, districts, and schools (under)serve EL students, all of which has been impacted by the pandemic. These disparities are the inevitable symptoms of a severely underfunded public education system that also expends its limited resources inequitably. While in a post-pandemic effort to offset some of the predicted learning losses and address some of these disparities the

---

[83] Teach for America: The Fight to Keep English Learners from Falling Through the Cracks. Jessica Frengi (2021). Retrieved from: https://www.teachforamerica.org/one-day/top-issues/the-fight-to-keep-english-learners-from-falling-through-the-cracks

[84] PBS Wisconsin. State Budget: English Language Learners. (2019). Retrieved from: https://pbswisconsin.org/news-item/state-budget-english-language-learners/

Elementary and Secondary School Emergency Relief (ESSER) funds allocated 190 billion USD to public education, researchers and practitioners have voiced concerns that these funds were insufficient, not well-targeted, and allocated without guidance on effective and productive investment areas.

Further, policymakers and administrators have been warning about an "ESSER spending cliff," as the timeline to allocate the funds expires in September 2024; for example, Roza and Silberstein (2023) report that the expiration of ESSER funds will leave states and districts staring down a massive fiscal cliff that equates to a single-year reduction in spending of over $1,000 per student.[85] Meanwhile, Peña and colleagues (2023) warn that absent "sustained education investments, the effects of the pandemic on children's educational progress will not wane" (p. 2). The results from this study corroborate this claim and indicate that the recovery, which may be sourced to these much needed, albeit lump-sum and fast-expiring funds, has thus far been small and inequitably distributed. The policy implications are simplistic, but unequivocal; more funding, and better targeted supports are needed to address the substantial, persistent, and growing disparities within the English learner student population, especially those between Hispanic and non-Hispanic identified ELs.

In addition to more and better-targeted support for these students, better data recording and reporting systems are needed for more accurate and rigorous studies. The high mobility of the EL population, missing data on important variables such as program types and various measures of SES, and lack of detailed and consistent demographic

[85] While the averages vary widely, this is approximately equal to a 10% reduction in the overall per pupil spending.

and educational data are some of the issues that force researchers to re-categorize or entirely exclude important variables or student subgroups from analyses. State educational agencies that are part of the WIDA Consortium are encouraged to continue pursuing rigorous data collection, reporting, and sharing mechanisms, enabling high-quality research to inform the theory and practice.

Contributions

This study makes several important theoretical and empirical contributions to the emergent literature on English Learners' education.

First and foremost, this study provides consistent and up-to-date evidence on the large and ongoing impact of the COVID-19 pandemic on EL's average proficiency. Worryingly, while there is a small upward trend in average proficiency recorded by students for the most recent, 2022-23 academic year (as compared to that in the previous year), the estimates show that this average increase is small, and disproportionately distributed. For example, Hispanic English Learners, also representing the largest and fastest-growing demographic group of students nationally, have experienced further increases in the already-sizeable disparities in average English proficiency, as compared to their non-Hispanic identified EL peers.

Second, it is the first study to provide precise and generalizable empirical evidence on English Learners' academic outcomes at this scale, scope, and granularity, while considering both temporal- and institutional-level variations in student proficiency, and for a large number of previously unexplored student-level variables, categories, and their intersections. Elaborating on this point further, the study uncovers persistent disparities within English Learner subgroups and documents that, perhaps expectedly, many of these disparities have been exacerbated by the COVID-19 pandemic. Differences in average proficiency outcomes across ethno-racial and intersectional subgroups, estimated leveraging large-scale linguistic assessment data and multilevel models, provide consistent evidence of differential outcomes across many subgroups.

Third, I present a theoretical *Framework of English Learner Intersectionality,* positioning socially-constructed English Learner status at the center of institutional (i.e., state-, district-, and school-level) factors that interact with ELs' overlapping identities in different and dynamic ways, shaping educational outcomes for socially constructed student subgroups (based on race, ethnicity, gender, ability, etc.). Examining the underlying data (which captures the entire universe of ACCESS data from WIDA states spanning pre- and post- COVID-19 periods) under the nuanced and multidimensional lens of *Intersectionality* surfaces disparities in educational outcomes of several EL student subgroups. In addition to documenting disparities between the average proficiency outcomes of subgroups categorized by race and ethnicity, the analysis quantifies relationships between important variables such as students' time in EL programming and interruptions to students' education, as well as non-trivial differences in the average outcomes of ELs by disability status, gender, migrant status, newcomer EL status, long-term status, and waived school supports, as well as interactions of several of these variables with Hispanic ethnicity. The latter provide important insights on how Hispanic ELs' educational outcomes differ from their non-Hispanic peers, depending on the additional intersections under inquiry. The additional interaction of all these variables with the COVID-19 binary variable, as already discussed in the first point of this list of contributions, provides many insights on how these individual- and institutional-level relationships have been impacted by the pandemic.

Fourth, this is the first study to examine EL outcomes while considering the multiply nested structure of these students' educational outcomes across time and within schools, districts, and states. The increasing flexibility of examined empirical models enables comparing reported parameter estimates across different model specifications and providing a rich description of statistical dependencies between individual level factors and EL proficiency. The final, five-level mixed-effects specification allows for a precise decomposition of variations in ELs proficiency, sourcing it to "fixed" individual-level factors and "random" student-, school-, district-, and state-level effects. The findings of this study confirm that the inclusion of these institutional contexts in the analysis inform the results in important ways. The changes in parameter estimates across multilevel model specifications signal differential ways in which states, districts, and schools have been impacted by the pandemic, and how specific EL subgroups have been differentially (under)served by these various levels of education. These estimates, while not causal, are very consistent and precise, and they can serve as a general reference for researchers examining relationships, differences, and variations within EL outcomes.

Fifth, this is the first study to quantify the impact of interruptions to EL students' formal education. Calculating the difference of students' age from the average age of their grade-level cohort, I include this variable in the empirical models as a proxy variable for

SLIFE. This provides a convenient way to evaluate differences in average proficiency associated with a one-year increase in the age difference.

Sixth, this is the first study to quantify differences in EL outcomes by *Migrant* status. Results of the analysis show that *Migrant* ELs are reporting lower average proficiency compared to ELs without *Migrant* identification. This difference, on average, is approximately equal to that of a year of interruption to students' education, as measured by *SLIFE*. These differences provide some food for thought for district and school administrators, educators, and parents about the important role of in- and out- of -school support systems, in turn moderated by students' socio-economic status, school attendance, and mobility.

Seventh, this is the first study to examine EL outcomes by *LIEP Waiver* status. ELs who refuse in-school language support services are reporting higher average scores than their EL peers who receive language support services at school. This difference is approximately equal to the difference estimated – in a final, eighth contribution by this study – between *Female* and *Male* English Learners.

Caveats, Limitations, and Future Research

There are several caveats and limitations in this work, leading to promising directions for future research.

First and foremost, there are several potentially important variables, such as program types (LIEPs), teachers, classrooms (peers), students' socio-economic status and native language, parents' education, degree of previous formal schooling, among others, the omission of which may under- or over-estimate the parameters describing the magnitude of differences between specific EL subgroups' outcomes. Further, the data do not include school, district, or state-level variables which also may be additionally predictive of EL proficiency, and further refine random-effects estimates. Therefore, while all effort has been made to control for all observed and unobserved heterogeneity through the inclusion of student, school, district, and state random effects, the estimated parameters, albeit very precise and consistent, should not be interpreted as causal and should be interpreted with caution. States, districts, and schools are encouraged to perform similar intersectional analyses for a more nuanced understanding of the local context. Inclusion of additional variables that are observed and measured across locales could be further informative in explaining variations in EL proficiency.

Second, while the estimated final specification is a very complex model with multiple fixed covariates, interactions, and four levels of nesting, the simplest error variance structure (random intercepts only) had to be applied to facilitate model convergence due to the large sample size and computing limitations. Meanwhile, the findings from this analysis signal that (a) the pandemic has also impacted the institutional levels of education in varied and different ways, and (b) there are substantial differences in average outcomes between several subgroups of EL students across and within these institutional levels. Therefore, a random-effects specification with a more flexible variance structure, for example including random slopes for (a) pre- and post-pandemic differences in outcomes and (b) for Hispanic ethnicity identification at the school-, district- and state-levels, would likely further inform the estimated statistical relationships. Similarly, a cross-nested structure of hierarchies that could allow for a more precise estimation of school random effects for students who move between schools across time is not feasible. Similarly, a cross-nested structure of hierarchies that could allow for a more precise estimation of school random effects for students who move between schools across time is not feasibleRegardless, these enhancements are left for future research, as WIDA is currently investigating the use of HPCs (High Performance Computers) which can perform quadrillions of calculations per second as compared to billions for regular computers that are a thousand times slower.

Third, while this study focuses on disparities by Hispanic ethnicity, many more intersections of student-level racial, gender, (dis)ability, and other categories, as well as interactions with some of the continuous covariates (e.g. Time as EL, or impact of SLIFE) could be considered. This is also left as an area for future research.

Fourth, the analysis is based exclusively on overall composite scale scores, which in turn are constructed by a weighted combination of students' scores in the four individual domains of Reading, Speaking, Listening, and Writing. Future research will examine EL outcomes in these domains separately, and jointly, for example in a *Seemingly Unrelated Regression* (SUREG) framework, for a more nuanced understanding of potential differences in academic language development across EL subgroups, as well as on how these differences have been impacted by the pandemic and the ensuing shifts in the delivery and modes of instruction.

Fifth, while this study provides unequivocal evidence about the "what," i.e., that the impact of the pandemic on ELs' education has been large, sustained, and differential, it cannot answer the "how," nor offer direct insights into the specific mechanisms driving the uncovered differential impacts and the uncovered disparities. Mixed-methods or qualitative inquiries may be better suited to address these questions in future research.

Sixth, due to the large scale and scope of the analysis, the discussion had to focus around the more important and consistent trends and differences, while there are many more detailed insights and nuances that can be gathered from the parameter estimates of 7 main and 18 auxiliary regression models, reported in Tables 4.1-4.4. Further, due to the emphasis in this analysis on disparities across ethno-racial intersections and differences by Hispanic interactions, relationships between other included demographic variables and EL proficiency were only briefly presented and described. Future analyses can more rigorously examine the reported disparities, differences, and pandemic-induced changes in thereof.

Despite these limitations, I remain hopeful that the methods and the findings and of this research, along with a number of questions left for future research, will prompt and promote further explorations of the systemic factors that continue to limit and restrict the access of marginalized student subgroups to more equitable educational opportunities. Researchers and practitioners are encouraged to use these results and compare them to estimates derived from their analyses, for a more nuanced and complete understanding of English Learners' education in more localized settings.

As the 2023-24 ACCESS administration is currently wrapping up, the results of this most recent proficiency assessment will be crucial in shedding further light on the impacts of the COVID-19 pandemic on EL's education. Meanwhile, the foundational data and variable management work, along with the statistical modeling performed in this study can serve as a springboard for future research using ACCESS data. Bolstered by additional and up-to-date assessment, demographic, and aggregate data and supported by higher computational power enabling more sophisticated modeling techniques, future research will examine the overall and differential impacts of the pandemic on ELs' education and evaluate whether the slight recovery that is likely due to ESSER funding efforts will be sustained going forward into school years of 2024 and 2025.

Conclusions

The COVID-19 pandemic has had a large and sustained negative impact on English Learners' education. This dissertation examined the extent of this impact, further focusing on identifying, quantifying, and documenting disparities within this very diverse subgroup of students. By examining EL's academic English proficiency — which largely determines students' English Learner status — this study considered various factors at the individual and institutional levels that have shaped these students' educational experiences and academic trajectories in American schools.

Leveraging population-level assessment and demographic data on students identified as ELs who take the ACCESS Online annual language proficiency assessment in WIDA Consortium states, this research presented evidence from regression models that account for the clustering of millions of students within thousands of schools and districts across WIDA states. The regression models examined individual-level variables like duration of EL status, newcomer and long-term designations, ethnicity, race, gender, disability, interrupted education, migrant status, and parental refusal of in-school language support services. Examining changes in mean and variance parameter estimates for individual- and institutional-level variables across various model specifications further informed nuanced relationships manifesting into differences and disparities in average English proficiency for several intersectional ethno-racial and other demographic subgroups.

The findings from multilevel regression models provided consistent evidence of persistent disparities in English proficiency between ELs identified by different ethno-racial subgroups, documenting the ongoing and varied impact of the pandemic on these students' academic outcomes. For instance, Hispanic students – a growing demographic of students already constituting the majority of the EL population in WIDA states and nationally – reported substantially lower proficiency levels compared to their non-Hispanic peers: a disparity that has been exacerbated by the pandemic. Newcomer ELs, on the other hand, scored substantially higher on their first ACCESS Online test after the pandemic—unless, again, they were also identified as Hispanic. Interpreting these disparities in terms of students' time in language instructional support programs, before the pandemic Hispanic English Learners (or all races) were, on average, about four months behind their not Hispanic peers; after the pandemic this disparity increased to about six months.

These differences in average proficiency across ethno-racial and other demographic subgroups were examined through the lens of Intersectionality (Crenshaw, 1991), helping illuminate how historical, political, and economic inequalities in educational opportunities have led to systemic disparities in academic outcomes. This approach helped reveal the varying effects of the pandemic on the education of English Learners, mostly intensifying existing inequalities for many vulnerable students.

Although there is some evidence of a modest post-COVID-19 recovery among certain EL subgroups, the findings underscore the urgent need for more, and better targeted supports for ELs, and especially for English Learners identified as Hispanic. Absent a significant and fundamental change in the education of these young language learners, the academic and career potential of many of these students will remain underrealized.

REFERENCES

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

Ackerman, D. J., & Tazi, Z. (2015). Enhancing young Hispanic dual language learners' achievement: exploring strategies and addressing challenges. *Policy Information Report, ETS Research Report No. RR-15-01. Princeton, NJ: Educational Testing Service.* http://dx.doi.org/10.1002/ets2.12045

Acosta, S., Garza, T., Hsu, H., Goodson, P., Padrón, Y., Goltz, H. H. & Johnson, A. (2019). The Accountability Culture: a Systematic Review of High-Stakes Testing and English Learners in the United States During No Child Left Behind. *Educational Psychology Review*, 32:327-352.

Adair, J. K. (2015). The impact of discrimination on the early schooling experiences of children from immigrant families.

Alliance for Excellent Education. Students of Color Caught in the Homework Gap (2021). Ret.: https://futureready.org/wp-ontent/uploads/2020/08/HomeworkGap_FINAL8.06.2020.pdf

Artiles, A. J. & Ortiz, A. (Eds.). (2002). English language learners with special needs: Identification, placement, and instruction. *Washington, DC: Center for Applied Linguistics*.

August, D., & Shanahan, T. (Eds.). (2006). Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth. Mahwah, NJ: Erlbaum

August, D., Shanahan, T., Pray, L., & Jiménez, R. T. (2009). Literacy and English-language learners: What researchers and policy makers should know. *Educational Researcher*, *38*(5), 380-385.

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2020). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193, 104345–104361.

Baker, B. D., Srikanth, A., Cotto Jr, R., & Green III, P. C. (2020). School Funding Disparities and the Plight of Latinx Children. *Education policy analysis archives*, 28(135), n135.

Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. TESOL Quarterly, 48(3), 480-506.

Bailey, A. L., & Kelly, K. R. (2013). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy*, *27*(5), 770-804.

Balart, P., & Oosterveen, M. (2019). Females show more sustained performance during test-taking than males. *Nature communications*, *10*(1), 3798.

Baltagi, B. H., & Liu, L. (2020). Forecasting with unbalanced panel data. *Journal of Forecasting*, *39*(5), 709-724.

Bartlett, C., Callahan, R., & Mavrogordato, M. (2024). Beyond a Transformative Approach and Deep Understanding: External Factors and Mid-Level Leaders' Policy Implementation to Expand Equity for English Learners. *Educational Administration Quarterly*, 0013161X241230296.

Baruch, B. (2023). The Impact of Remote Learning on the English Acquisition of K12 Multilingual Learners: Evidence from Public Schools in Delaware (Doctoral dissertation, Notre Dame of Maryland University).

Bauer, G. R., Churchill, S. M., Mahendran, M., Walwyn, C., Lizotte, D., & Villa-Rueda, A. A. (2021). Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM-population health*, *14*, 100798.

Biden, J. (2021). Statement by President Joe Biden on Armenian Remembrance Day. Retrieved: https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/24/statement-by-president-joe-biden-on-armenian-remembrance-day/

Bilingual Education Act. (1968). Public Law No. (90-247), 81 Stat.

Boals, T., Kenyon, D. M., Blair, A., Cranley, M. E., Wilmes, C., & Wright, L. J. (2015). Transformation in K–12 English Language Proficiency Assessment: Changing Contexts, Changing Constructs. Review of Research in Education, 39(1), 122-164. https://doi-org.ezproxy.library.wisc.edu/10.3102/0091732X14556072

Bond, K. M. (2020). Navigating Language Policy in State Education Agencies: English Learner Reclassification in WIDA States. *Available from ProQuest Dissertations & Theses Global; Publicly Available Content Database. (2423100154).*

Bonilla-Silva, E. (2015). The structure of racism in color-blind, “post-racial” America. *American Behavioral Scientist*, *59*(11), 1358-1376.

Bonilla-Silva, E. (1996). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, 62(3), 465-480.

Boone, J. H. (2013). ¡Ya Me Fui! When English Learners Consider Leaving School. *Education and Urban Society*, 45(4), 415-439. https://doi-org.ezproxy.library.wisc.edu/10.1177/0013124511413121

Bowleg, L. (2012) The problem with the phrase women and minorities: Intersectionality—an important theoretical framework for public health. *American Journal of Public Health*, 102 (7) (2012), pp. 1267-1273, 10.2105/AJPH.2012.300750.

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

Brooks, M. D. (2017). “She doesn’t have the basic understanding of a language”: Using spelling research to challenge deficit conceptualizations of adolescent bilinguals. *Journal of Literacy Research*, 49(3), 342-370.

Bryner, L. (2021). The teacher shortage in the United States. *Education and Society*, *39*(1), 69-80.

Buenrostro, M., & Maxwell-Jolly, J. (2021). Renewing our promise: Research and recommendations to support California’s long-term English learners. *Californians Together*.

Burr, E., Haas, E., & Ferriere, K. (2015). Identifying and Supporting English Learner Students with Learning Disabilities: Key Issues in the Literature and State Practice. *REL 2015-086. Regional Educational Laboratory West*.

Butler, Y. G., & Le, V. N. (2018). A longitudinal investigation of parental social-economic status (SES) and young students’ learning of English as a foreign language. *System*, *73*, 4-15.

Immigrant Optimism and Ever-English Learners’ Journeys from Kindergarten to College in Texas

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. The Future of Children, 21(1), 103–127. https://doi.org/10.1353/foc.2011.0007.

Callahan, R. M. (2013). The English learner dropout dilemma: Multiple risks and multiple resources. *Santa Barbara: California Drop out Research Project.*

Callahan, R. M., & Muller, C. (2013). *Coming of political age: American schools and the civic development of immigrant youth*. Russell Sage Foundation. Callahan, R. M., & Muller, C. (2013). *Coming of political age: American schools and the civic development of immigrant youth*. Russell Sage Foundation.

Callahan, R. M., Jiang, L., & Núñez, A.-M. (2023). EL Policy and Immigrant Politics: State and Federal Policy and Ever-EL Students' Postsecondary Pathways. *Educational Policy*, 37(5), 1467-1505. https://doi.org/10.1177/08959048221103802

Callahan R. M., Hopkins M. (2017). Using ESSA to improve secondary English learners' opportunities to learn through course taking. *Journal of School Leadership*, 27(5), 755–766.

Callahan, R. M., & Humphries, M. H. (2016). Undermatched? School-based linguistic status, college going, and the immigrant advantage. *American Educational Research Journal*, *53*(2), 263-295.

Callahan R., Shifrer D. (2016). Equitable access for secondary English learner students course taking as evidence of EL program effectiveness. *Educational Administration Quarterly*, 53, 463–496.

Callahan R. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal*, 42(2), 305–328

Callahan R., Wilkinson L., Muller C. (2010). Academic achievement and course taking among language minority youth in U.S. schools: Effects of ESL placement. *Educational Evaluation and Policy Analysis*, 32(1), 84–117.

Carhill, A., Suárez-Orozco, C., & Páez, M. (2008). Explaining English language proficiency among adolescent immigrant students. *American Educational Research Journal*, 45, 1155–1179. doi:10.3102/0002831208321443

Carnoy M., Garcia E. (2017). Five key trends in U.S. Student performance. *Economic Policy Institute Report*. http://www.epi.org/publication/five-key-trends-in-u-s-student-performance-progress-by-blacks-and-hispanics-the-takeoff-of-asians-the-stall-of-non-english-speakers-the-persistence-of-socioeconomic-gaps-and-the-damaging-effect/

Castañeda v. Pickard. (5th Cir. 1981). 648 F.2d 989.

Castillo, W., & Gillborn, D. (2022). How to "QuantCrit:" Practices and questions for education data researchers and users. *Manuscript under review*.

Cho, S., Crenshaw, K. W., & McCall, L. (2013). Toward a field of intersectionality studies: Theory, applications, and praxis. Signs: *Journal of women in culture and society*, 38(4), 785-810.

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English Learner Reclassification Policy Effects Across Districts. *American Educational Research Journal, 54(1_suppl)*, 255S-278S.

Clark-Gareca, B., Short, D., Lukes, M., & Sharp-Ross, M. (2020). Long-term English learners: Current research, policy, and practice. *TESOL J. 2020*; 11:e452.

Clayton, A. (2021). *Bernoulli's fallacy: Statistical illogic and the crisis of modern science*. Columbia University Press.

Conger, D., Long, M. C., & Iatarola, P. (2009). Explaining race, poverty, and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management*, *28*(4), 555-576.

Cook, H. G., Boals, T., & Lundberg, T. (2011). Academic achievement for English learners: What can we reasonably expect?. *Phi Delta Kappan*, *93*(3), 66-69.

Cook, H. G., & Linquanti, R. (2015). Strengthening Policies and Practices for the Initial Classification of English Learners: Insights from a National Working Session. *Council of Chief State School Officers*.

Cornell, Stephen E. (Stephen Ellicott), 1948-. (2007). Ethnicity and race: making identities in a changing world. *Thousand Oaks, Calif. :Pine Forge Press, an Imprint of Sage Publication*

Coulter, K. (2016). Effectiveness of WIDA English language development standards meeting proficiency for English language learners (Order No. 10243679). *Available from ProQuest Dissertations & Theses Global. (1865652058).*

Cornwell C, Mustard DB, & Van Parys, J. Noncognitive skills and the gender disparities in test scores and teacher assessments: evidence from primary school. J. Hum. Resour. 2013;48:236–264.

Craver, E. (2009). On the Boundary of White: The Cartozian Naturalization Case and the Armenians, 1923-1925. *Journal of American Ethnic History*, 28(2), 30–56. http://www.jstor.org/stable/40543386

Crenshaw, K. W. (2013). Mapping the margins: Intersectionality, identity politics, and violence against women of color. In *The public nature of private violence* (pp. 93-118). Routledge.

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43(6), 1241–1299. https://doi.org/10.2307/1229039

Cummins, J. (2021b). Rethinking the Education of Multilingual Learners. *Bristol, UK: Multilingual Matters.*

Daniels, J., & Westerlund, R. (2018). WIDA focus on: Scaffolding learning for multilingual students in math. *Wisconsin Center for Education Research*.

Darling-Hammond, L. (2007) Race, inequality and educational accountability: the irony of 'No Child Left Behind'. *Race Ethnicity and Education*, 10(3):245-260.

DeCapua, A., Smathers, S., & Tang, L. (2009). Meeting the Needs of Students with Limited or Interrupted Schooling. Ann Arbor, MI: *The University of Michigan Press.*

DiCerbo, P. A. (2001). Why migrant education matters. *National Clearinghouse for Bilingual Education Issue Brief*, *8*.

DiCerbo, P. A., Anstrom, K. A., Baker, L. L., & Rivera, C. (2014). A review of the literature on teaching academic English to English language learners. *Review of educational research*, 84(3), 446-482.

Delgado, Richard (1995). Critical Race Theory: *The Cutting Edge*

Dorn, E., Hancock, B., Sarakatsannis, J. (2020). COVID-19 and Learning Loss—Disparities Grow and Students Need Help. *MCKINSEY & CO*.

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, *98*(1), 198.

Dynarski, S. (2018, January 19). Online courses are harming the students who need the most help. *The New York Times*. https://www.nytimes.com/2018/01/19/business/online-coursesareharming-the-students-who-need-the-most-help.html

Education Trust. (2020). 10 questions for equity advocates to ask about distance learning. *The Education Trust*. https://edtrust.org/resource/10-questions-for-equity-advocates-to-ask-aboutdistance-learning

Every Student Succeeds Act. (2015). Public Law (114-95).

Estrada, P. & Wang, H. (2018). Making English Learner Reclassification to Fluent English Proficient Attainable or Elusive: When Meeting Criteria Is and Is Not Enough. *American Educational Research Journal*, 55(2):207-242.

Faulkner-Bond, M. & Sireci, S. (2015). Validity Issues in Assessing Linguistic Minorities. *International Journal of Testing*. 15. 10.1080/15305058.2014.974763.

Ferron, J. and Rendina-Gobioff, G. (2005). Interrupted Time Series Design*. In Encyclopedia of Statistics in Behavioral Science* (eds B.S. Everitt and D.C. Howell). https://doi-org.ezproxy.library.wisc.edu/10.1002/0470013192.bsa312

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 & Education*, *14*(2), 113-132.

Fox, J., & Fairbairn, S. (2011). Test review: ACCESS for ELLs®. *Language Testing*, 28(3), 425-431. https://doi-org.ezproxy.library.wisc.edu/10.1177/0265532211404195

Frengi, J. (2021). The Fight to Keep English Learners from Falling Through the Cracks. *Teach for America*. https://www.teachforamerica.org/one-day/top-issues/the-fight-to-keep-english-learners-from-falling-through-the-cracks

Fry, R. (2008). The role of schools in the English language learner achievement gap. *Washington, DC: Pew Hispanic Center*.

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

Gamoran A. (2010). Tracking and inequality: New directions for research and practice*. In The Routledge international handbook of the sociology of education* (pp. 213–228). Abingdon, UK: Routledge.

Gándara, P. C., & Contreras, F. (2009). The Latino Education Crisis: The Consequences of Failed Social Policies. *Cambridge, MA: Harvard University Press*.

Gándara, P. & Orfield, G. (2012). Segregating Arizona's English learners: A return to the "Mexican room." *Teachers College Record*, 114(9):1-27.

Gándara, P., & Rumberger, R. W. (2009). Immigration, language, and education: How does language policy structure opportunity? *Teachers College Record,* 111(3), 750-782.

Gándara P., Rumberger R., Maxwell-Jolly J., Callahan R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, 11(36), 1–54.

Garbe, A., Ogurlu, U., Logan, N., & Cook, P. (2020). Parents' experiences with remote education during COVID-19 school closures. *American Journal of Qualitative Research*, 4(3), 45–65. https://doi.org/10.29333/ajqr/8471

García, O., & Kleifgen, J. A. (2020). Translanguaging and literacies. *Reading research quarterly*, 55(4), 553-571.

Garver, R. (2020). How Harmful Is Segregation? English Learners' Conditions For Learning In Segregated Classrooms*. Leadership and Policy in Schools*, 19(1):123-140.

Geeslin, K. L., & Long, A. Y. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context*. Routledge.

Gelatt, J., Lacarte, V., & Rodriguez, J. (2022). A Profile of Low-Income Immigrants in the United States. *Washington, DC: Migration Policy Institute*.

Gottlieb, M., Cranley, M. E., & Cammilleri, A. (2007). Understanding the WIDA English language proficiency standards: *A resource guide. Madison, WI: WIDA Consortium*.

Gottlieb, M., & Kenyon, D. (2006). History of research intensive test development: The Bridge Study between Tests of English Language Proficiency and ACCESS for ELLs. *WIDA Consortium Technical Report No.* 2.

Hamayan, E., Marler, B., Sánchez-Lopez, C. & Damico, J. (2013). Special Education Considerations for English Language Learners. *Second Edition. Carlson*. Philadelphia, PA.

Hamilton, L. S., Kaufman, J. H., & Diliberti, M. (2020). Teaching and leading through a pandemic: Key findings from the American Educator Panels Spring 2020 COVID-19 surveys. https://www.rand.org/pubs/research_reports/RRA168-2.html

Hancock, A. M. (2013). Empirical intersectionality: A tale of two approaches, 3 UC Irvine L. Rev. 259.

Hancock, A. M. (2007). When multiplication doesn't equal quick addition: Examining intersectionality as a research paradigm. *Perspectives on politics*, *5*(1), 63-79.

Hankivsky, O., & Cormier, R. (2019). Intersectionality and public policy: Some lessons from existing models. *The Palgrave handbook of intersectionality in public policy*, 69-93.

Hankivsky, O., & Cormier, R. (2011). Intersectionality and Public Policy: Some Lessons from Existing Models. *Political Research Quarterly*, 64(1), 217–229. http://www.jstor.org/stable/41058335

Harklau, L. (2016). Bureaucratic dysfunctions in the education of Latino immigrant youth. *American Journal of Education*, *122*(4), 601-627.

Heissel, J. (2016). The relative benefits of live versus online delivery: Evidence from virtual algebra I in North Carolina. *Economics of Education Review*, *53*, 99-115.

Hill, L., Weston M., & Hayes J. (2014). Reclassification of English Learner Students in California. *San Francisco: Public Policy Institute of California*.

Hopkins, M., Lowenhaupt, R. & Sweet, T. M. (2015). Organizing English Learner Instruction in New Immigrant Destinations: District Infrastructure and Subject-Specific School Practice. *American Educational Research Journal*, 52(3):408-439.

Hopkins, M., Thompson, K. D., Linquanti, R., Hakuta, K., & August, D. (2013). Fully accounting for English learner performance: A key issue in ESEA reauthorization. Educational Researcher, 42, 101–108

Huck C., Zhang J. (2021). Effects of the COVID-19 Pandemic on K–12 education: A systematic literature review. *New Waves-Educational Research and Development Journal*, 24(1), 53–84.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological bulletin*, *104*(1), 53.

Individuals with Disabilities Education Act (1975). Public Law No. (101-476), 104 Stat. 1142.

Jiménez-Castellanos, O. H., & García, D. (2017). School expenditures and academic achievement differences between high-ELL-performing and low-ELL-performing high schools. *Bilingual Research Journal*, 40(3), 318-330. https://doi.org/10.1080/15235882.2017.1342717

Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education review*, *21*(6), 589-598.

Johnson, A. (2022). Achievement and Growth for English Learners*. Educational Policy*.

Kangas, S. E. N. (2014). When special education trumps ESL: An investigation of service delivery for ELLs with disabilities. *Critical Inquiry in Language Studies*, 11(4):273-306.

Kangas, S. E. N. (2017). A cycle of fragmentation in an inclusive age: The case of English learners with disabilities. *Teaching and Teacher Education*, 66:261-272.

Kangas, S. E. N. (2018). Breaking one law to uphold another: Service provision for English learners with disabilities. *TESOL Quarterly*, 52(4):877-910.

Kangas, S. E. N., & Cook, M. (2020). Academic Tracking of English Learners With Disabilities in Middle School. *American Educational Research Journal*, 57(6):2415-2449.

Kanno, Y., & Cromley, J. G. (2013). English language learners' access to and attainment in postsecondary education. *Tesol Quarterly*, *47*(1), 89-121.

Kanno, Y., & Kangas, S. E. N. (2014). "I'm Not Going to Be, Like, for the AP": English Language Learners' Limited Access to Advanced College-Preparatory Courses in High School*. American Educational Research Journal,* 51(5), 848-878. https://doi-org.ezproxy.library.wisc.edu/10.3102/0002831214544716

Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual review of sociology*, *29*(1), 417-442.

Kenyon, D. M., Ryu, J. R., & MacGregor, D. (2007). Setting Grade Level Cut Scores for ACCESS for ELLs®. *WIDA Consortium Technical Report No. 4*.

Kibler, A. K., Karam, F. J., Ehrlich, V. A., Bergey, R., Wang, C. & Elreda, L. M. (2018). Who are 'long-term English learners'? Using classroom interactions to deconstruct a manufactured learner label. *Applied Linguistics*, 39(5):741-765.

Kieffer, M. J., & Parker, C. E. (2016). Patterns of English Learner Student Reclassification in New York City Public Schools. *Regional Educational Laboratory Northeast & Islands.* REL 2017-200.

Kieffer, Michael J., and Carolyn E. Parker. 2017. Graduation Outcomes of Students Who Entered New York City Public Schools in Grade 5 or 6 As English learner Students. *Regional Educational Laboratory Northeast & Islands.* REL 2017-237.

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, *39*(6), 484-486.

Kim, A., Yumsek, M., Chapman, M. & Cook, H. G. (2019). Investigating K-12 English Learners' Use of Universal Tools Embedded in Online Language Assessments (*WIDA Technical Report No. TR-2019-2*). Madison, WI: WIDA at the Wisconsin Center for Education Research.

Kim, A. A., Molle, M., Kemp, J., & Cook, H. G. (2018). Examination of identification and placement decisions made for K–12 English learners (WCER Working Paper No. 2018-12). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/working-papers

Kim, S., Chang, M., Singh, K., & Allen, K. R. (2015). Patterns and Factors of High School Dropout Risks of Racial and Linguistic Groups. *Journal of Education for Students Placed at Risk (JESPAR)*, 20(4), 336–351. https://doi-org.ezproxy.library.wisc.edu/10.1080/10824669.2015.1047019

Kim, Y. K., Curby, T. W., & Winsler, A. (2014, October 13). Child, Family, and School Characteristics Related to English Proficiency Development Among Low-Income, Dual Language Learners. *Developmental Psychology. Advance online publication*. http://dx.doi.org/10.1037/a0038050

Kim, W. G. & García, S. B. (2014). Long-Term English Language Learners' perceptions of their language and academic learning experiences. *Remedial and Special Education*, 35(5), 300-312.

Ladson-Billings, G., & Tate, W. F. (1995). Toward a Critical Race Theory of Education. *Teachers College Record*, 97(1), 47-68. https://doi-org.ezproxy.library.wisc.edu/10.1177/016146819509700104

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher*, 35(7):3-12.

Lau v. Nichols. (1974). 414 US 563

Le, B., Black, K. E., Carlson, C., Miciak, J., Romano, L., Francis, D., & Kieffer, M. J. (2024). Ever English Learner 4-Year Graduation: Toward an Intersectional Approach. *Educational Researcher*, 0(0). https://doi-org.ezproxy.library.wisc.edu/10.3102/0013189X241246747

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

Linquanti, R. & Cook, G. (2013). Toward a "common definition of English learner": Guidance for states and state assessment consortia in defining and addressing policy and technical issues and options. *Washington DC: Council of Chief State School Officers*.

Liu, K. K., Thurlow, M. L., Press, A. M., & Lickteig, O. (2018). A review of the literature on measuring English language proficiency progress of English learners with disabilities and English learners (NCEO Report 408). *Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes*.

Lillie, K. E., Markos, A., Arias, M. B., & Wiley, T. G. (2012). Separate and not equal: The implementation of structured English immersion in Arizona's classrooms. *Teachers College Record*, *114*(9), 1-33.

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S. & Reber, S. (2017). Descriptive analysis in education: A guide for researchers (No. NCEE 2017-4023). *Washington, DC: National Center for Education Evaluation and Regional Assistance*

Lopez, A. A., Pooler, E., & Linquanti, R. (2016). Key issues and opportunities in the initial identification and classification of English learners. *ETS Research Report Series*, *2016*(1), 1-10.

López, L. M., & Foster, M. E. (2021). Examining heterogeneity among Latino dual language learners' school readiness profiles of English and Spanish at the end of head start. *Journal of Applied Developmental Psychology*, *73*, 101239.

López, N., Erwin, C., Binder, M., & Chavez, M. J. (2018). Making the invisible visible: advancing quantitative methods in higher education using critical race theory and intersectionality. *Race Ethnicity and Education*, 21(2), 180–207. https://doi-org.ezproxy.library.wisc.edu/10.1080/13613324.2017.1375185

MacGregor, D., Yen, S. J., & Yu, X. (2022). Using Multistage Testing to Enhance Measurement of an English Language Proficiency Test. *Language Assessment Quarterly*, 19(1), 54–75. https://doi-org.ezproxy.library.wisc.edu/10.1080/15434303.2021.1988953

Marshall, D. T., Shannon, D. M., & Love, S. M. (2020). How teachers experienced the COVID-19 transition to remote instruction. *Phi Delta Kappan*, *102*(3), 46-50.

Mason-Williams, L., Rosenberg, M., Kimmel, L., & Sindelar, P. (2020). Addressing Shortages of Educators in an Uncertain COVID-19 Landscape: Viewing Teacher Candidates as Assets. *Center on Great Teachers and Leaders*.

Martinez, J. A., & Broemmel, A. D. (2021). Pencils down: Educators respond to the uncertainty amidst COVID-19 school closures. *International Studies in Educational Administration*, 49(1), 109-132.

Mavrogordato, M., Callahan, R., & Bartlett, C. (2022). Mid-level leaders as key policy interpreters: state and local leaders' perspectives on leveraging Castañeda to expand equity for English learner students. *Language policy*, *21*(3), 331-355.

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of women in culture and society*, *30*(3), 1771-1800.

Menken, K. (2008). English learners left behind: Standardized testing as language policy (Vol. 65). *Cleveden, England: Multilingual Matters*.

Menken, K. & Kleyn, T. (2010). The long-term impact of subtractive schooling in the educational experiences of secondary English language learners. *International Journal of Bilingual Education and Bilingualism*, 13(4):399-417

Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on "long-term English language learners": Characteristics and prior schooling experiences of an invisible population. *International Multilingual Research Journal*, *6*(2), 121-142.

Milner, H. R. (2012). Beyond a Test Score: Explaining Opportunity Gaps in Educational Practice. *Journal of Black Studies*, 43(6):693-718.

Misra, J. (2021). The intersectionality of precarity.

Molle, D., & Lee, N. (2017). Opportunities for academic language and literacy development for emergent bilingual students during group work. *International Journal of Bilingual Education and Bilingualism*, *20*(5), 584-601.

Molle, D., Sato, E., Boals, T., Hedgspeth, C. A. (Eds.) (2015). Multilingual learners and academic literacies: Sociocultural contexts of literacy development in adolescents. *New York: Routledge*.

Murphy, M., & Johnson, A. (2023). Dual Identification? The Effects of English Learner (EL) Status on Subsequent Special Education (SPED) Placement in an Equity-Focused District. *Educational Evaluation and Policy Analysis*, 45(2), 311-335.

National Center for Education Statistics. (2023). English Learners in Public Schools. Condition of Education. U.S. Department of Education, Institute of Education Sciences. Retrieved [4/18/24], from https://nces.ed.gov/programs/coe/indicator/cgf.

National Council of Teachers of English [NCTE]. (2008). English language learners: A policy research brief produced by the National Council of Teachers of English.

National Academies of Sciences, Engineering, and Medicine (2017). Promoting the Educational Success of Children and Youth Learning English: Promising Futures. *Washington, DC: The National Academies Press.* https://doi.org/10.17226/24677

New York State Education Department (2011). Guidelines for Educating Limited English Proficient Students with Interrupted Formal Education. 12-13. *Albany, NY: State Education Department*.

Nordmeyer, J., Boals, T., MacDonald, R. & Westerlund, R. (2021). What does equity really mean for multilingual learners? *ACSD Newsletter*. https://www.ascd.org/el/articles/what-does-equity-really-mean-for-multilingual-learners

Nowiki, M. (2020) Distance Learning: Challenges Providing Services to K-12 English Learners and Students with Disabilities during COVID-19. *United States Government Accountability Office. Report to Congressional Committee*. Retrieved from: https://www.gao.gov/assets/gao-21-43.pdf

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications*, 9(1), 3777.

Olivares, M. (2022). Funding Allocations and Human Resources: Identifying Opportunity Gaps for English Language Learners in "Low-Achieving" and "High-Achieving" Districts in California. *Journal of Education Human Resources. 23*.

Olsen, L. (2010). Reparable harm: Fulfilling the unkept promise of educational opportunity for California's long-term English learners. *Long Beach, CA: Californians Together*.

Olsen, L. (2014). Meeting the unique needs of long term English language learners: A guide for educators. *Washington, DC: National Education Association*.

Orfield, G., Ee, J., Frankenberg, E., & Siegel-Hawley, G. (2016). Brown at 62: School segregation by race, poverty and state.

Poole, G. A. (2024) [forthcoming]. Teacher Attrition and Opportunity Gaps for Multilingual Learners. (Doctoral dissertation, University of Madison-Wisconsin)

Poole, G. A., & Sahakyan, N. (2024). Examining English Learner Testing, Proficiency, and Growth: Continued Trends since the COVID-19 Pandemic. (*WIDA Research Report No. RR-2024-1*). Wisconsin Center for Education Research.

Poole, G., & Sahakyan N. (2023). Examining English Learner Testing, Proficiency, and Growth: Before, During, and "After" the COVID-19 Pandemic. (*WIDA Research Report No. RR-2023-1*). Wisconsin Center for Education Research.

Porter, T., Cook, H. G., Sahakyan, N. (2019). Less Than Four Domains: Creating an Overall Composite Score for English Learners with Individualized Education Plans (*WIDA Technical Report No. TR-2019-3*). Madison, WI: WIDA at the Wisconsin Center for Education Research.

Quinn, D. M., & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher*, *44*(6), 336-346.

Quiros, L., & Dawson, B. A. (2013). The Color Paradigm: The Impact of Colorism on the Racial Identity and Identification of Latinas. Journal of Human Behavior in the Social Environment, 23(3), 287–297. https://doi.org/10.1080/10911359.2012.740342

Rabe-Hesketh, S. & Skrondal, A. (2021) 'Multilevel and Longitudinal Modeling Using Stata Volume I: Continuous Responses'. *4th ed. (Texas, USA: Stata Press, 2012)*

Ragin, C. (2008). Redesigning Social Sciences. Fuzzy Sets and Beyond.

Ream, R. K., Ryan, S. & Yang, T. (2017). Reframing the problematic achievement gap narrative to structure educational success. In G. Q. Conchas, M. A. Gottfried, B. M. Hinga, & L. Oseguera (Eds.), Educational policy goes to school: Case studies on the limitations and possibilities of educational innovation (pp. 247-265*). New York: Routledge*.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American educational research journal*, *46*(3), 853-891.

Roberts, D. (2011). *Fatal invention: How science, politics, and big business re-create race in the twenty-first century*. New Press/ORIM.

Rodriguez, D., Carrasquillo, A., Garcia, E., & Howitt, D. (2020). Factors that challenge English learners and increase their dropout rates: Recommendations from the field. *International Journal of Bilingual Education and Bilingualism*, 1–17. https://doi-org.ezproxy.library.wisc.edu/10.1080/13670050.2020.1722059

Rosa, Jonathan, Looking like a Language, Sounding like a Race: Raciolinguistic Ideologies and the Learning of Latinidad, *Oxf Studies in Anthropology of Language* (New York, 2019; online edn, Oxford Academic, 21 Feb. 2019), https://doi.org/10.1093/oso/9780190634728.001.0001, accessed 28 Apr. 2024.

Rosenberg, D., & Anderson, T. (2021). Teacher Turnover before, during, & after COVID. *Education resource strategies*.

Ruiz-de-Velasco J., Fix M. (2000). Overlooked and underserved: Immigrant students in U.S. secondary schools. *Washington, DC: Urban Institute*

Rumberger, R. W., & Gándara, P. (2004). Seeking Equity in the Education of California's English Learners. *Teachers College Record*, 106(10), 2032–2056. https://doi.org/10.1111/j.1467-9620.2004.00426.x

Rumberger, R. W., & Gándara, P. (2000). The schooling of English learners.

Russell, M. (2023). Systemic Racism and Educational Measurement: Confronting Injustice in Testing, Assessment, and Beyond (1st ed.). *Routledge*. https://doi.org/10.4324/9781003228141

Sahakyan, N. (2015). Evaluating teacher effectiveness using ACCESS for ELLs (WIDA Research Brief). Madison, WI: Board of Regents of the University of Wisconsin System. Retrieved from: www.wida.us/downloadLibrary.aspx

Sahakyan, N. (2020). Generating Imputed* Overall Composite Scale Scores for English Learners with Disabilities Who Are Missing Domain Scores in the ACCESS for ELLs Assessment. WIDA Technical Report TR-2020-1. Retrieved from: https://wida.wisc.edu/sites/default/files/resource/technical-report-generating-imputed-overall-composite-scale-scores-english-learners-disabilities.pdf

Sahakyan, N., & Poole, G. (2023). “Every” Student Succeeds? Academic Trends at the Intersection of (Long-Term) English Learner and IEP Status. Journal of Education for Students Placed at Risk (JESPAR), 28(1), 69–96. https://doi-org.ezproxy.library.wisc.edu/10.1080/10824669.2022.2123328

Sahakyan, N., & Poole, G.A. (2021). Examining Growth at the Intersection of IEP and (Long Term) EL Status. *WIDA Research Report No. RR-2021-2. Madison, WI:* WIDA at the Wisconsin Center for Education Research.

Sahakyan, N., & Ryan, S. (2018). Long-term English learners across 15 WIDA states: A research brief. (*WIDA Research Brief No. RB-2018-1). Madison, WI:* WIDA at the Wisconsin Center for Education Research.

Santibañez, L., & Guarino, C. M. (2021). The Effects of Absenteeism on Academic and Social-Emotional Outcomes: Lessons for COVID-19. *Educational Researcher,* 50(6), 392–400. https://doi.org/10.3102/0013189X21994488

Sato, E., & Thompson, K. D. (2020). Standards-based K–12 English language proficiency assessments in the United States: Current policies and practices. In Assessing English Language Proficiency in US K–12 Schools (pp. 19-35). *Routledge*

Saunders, W. M. & Marcelletti, D. J. (2013). The Gap That Can't Go Away: The Catch-22 of Reclassification in Monitoring the Progress of English Learners. *Educational Evaluation and Policy Analysis*, 35(2):139-156. http://www.jstor.org/stable/43773425

Schaefer, M. B., Abrams, S. S., Kurpis, M., Abrams, M., & Abrams, C. (2020). " Making the Unusual Usual:" Students' Perspectives and Experiences of Learning at Home during the COVID-19 Pandemic. *Middle Grades Review*, *6*(2), n2.

Schissel, J. L., & Kangas, S. E. (2018). Reclassification of emergent bilinguals with disabilities: The intersectionality of improbabilities. *Language policy*, 17(4):567-589. https://doi.org/10.1007/s10993-018-9476-4.

Schmid, C. L. (2001). Educational achievement, language-minority students, and the new second generation. *Sociology of Education*, 71-87.

Scott, N. A., & Siltanen, J. (2017). Intersectionality and quantitative methods: Assessing regression from a feminist perspective. *International Journal of Social Research Methodology*, *20*(4), 373-385.

Shafer, G. (2001). Standard English and the migrant community. *The English Journal*, *90*(4), 37-43.

Shafer Willner, L., & Monroe, M. (2016). The WIDA accessibility and accommodations framework: Considerations influencing the framework development. Madison, Wisconsin: WIDA.

Shifrer, D., Muller, C. & Callahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. *Journal of learning disabilities*, 44(3):246-257.

Shin, N. (2020). Stuck in the middle: examination of long-term English learners. International multilingual research journal, 14(3):181-205

Short, Deborah J., & Boyson, Beverly A. (2012). Helping Newcomer Students Succeed in Secondary Schools and Beyond. *Washington, D.C.: Center for Applied Linguistics*.

Silberstein, K., & Roza, M. (2023). The ESSER fiscal cliff will have serious implications for student equity.

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(2):220-252.

Slama, R., Molefe, A., Gerdeman, D., Herrera, A., de los Reyes, I. B., August, D. & Cavazos, L. (2017). Time to proficiency for Hispanic English learner students in Texas. *American Institutes for Research*.

Smiley, C., Browne, A., & Battle, J. (2023). Latinx Parents, Latinx Students, and In-School Suspension: A Quantitative Study of School Discipline. *Journal of Latinos and Education, 22(4), 1622–1635*. https://doi.org/10.1080/15348431.2022.2026223

Solórzano, R. (2008). High Stakes Testing: Issues, Implications, and Remedies for English Language Learners. *Review of Educational Research*, 78(2):260-329.

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

Stevens, J. J. & Schulte, A. C. (2017). The interaction of learning disability status and student demographic characteristics on mathematics growth. *Journal of Learning Disabilities*, 50(3):261-274.

Sugarman, J. (2019). Legal protections for K-12 English learner and immigrant-background students. *Migration Policy Institute*, *3*.

Sugarman, J. & Geary, C. (2018). English learners in Minnesota: Demographics, Outcomes, and State Accountability Policies*. Washington, DC: Migration Policy Institute*.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, 77(3):317-334.

Tabron, L. A. & Thomas, A. K. (2023). Deeper than Wordplay: A Systematic Review of Critical Quantitative Approaches in Education Research. *Review of Educational Research*, 93(5):756-786.

Thompson, K. D. (2017). English Learners' Time to Reclassification: An Analysis. *Educational Policy*, 31(3):330-363.

Thompson, K. D. (2015). Questioning the Long-Term English Learner Label: How Categorization Can Blind Us to Students' Abilities. *Teachers College Record 117* (12): 1–50. doi:10.1177/016146811511701203.

Thompson, K. D., Umansky, I. M. & Rew, W. J. (2022) Improving Understanding of English Learner Education Through an Expanded Analytic Framework. *Educational Policy*, 34(1).

Tindal, G. & Anderson, D. (2019). Changes in status and performance over time for students with specific learning disabilities*. Learning Disability Quarterly*, 42(1):3-16. doi:10.1177/0731948718806660

Tyson, N. deGrasse (2021). "Neil deGrasse Tyson Explains Albedo." *Conversations with Chuck Nice.* Retrieved from: https://www.youtube.com/watch?v=pJ0GQYiBg_U&t=141s

Umansky, I. M. (2016). Leveled and exclusionary tracking: English learners' access to academic content in middle school. *American Educational Research Journal*, 53:1792-1833.

Umansky, I. M., & Avelar, J. D. (2023). Canaried in the Coal Mine: What the Experiences and Outcomes of Students Considered Long-term English Learners Teach us About Pitfalls in English Learner Education…and What We Can Do About It. *Journal of Education for Students Placed at Risk (JESPAR)*, 28(1), 122–147. https://doi-org.ezproxy.library.wisc.edu/10.1080/10824669.2022.2123326

Umansky I. M., Callahan R., Lee J. C. (2020). Making the invisible visible: Identifying and interrogating ethnic differences in English learner reclassification. *American Journal of Education*, 126(3), 335–388.

Umansky, I. M., & Porter, L. (2020). State English Learner Education Policy: A Conceptual Framework to Guide Comprehensive Policy Action. *Education Policy Analysis Archives*, *28*(17), n17.

Umansky, I. M., Thompson, K. D. & Díaz, G. (2017). Using an ever-English learner framework to examine disproportionality in special education. *Exceptional Children*, 84(1):76-96.

Umansky, I. M., Thompson, K. D., Soland, J. & Kibler, A. K. (2022). Understanding Newcomer English Learner Students' English Language Development: Comparisons and Predictors. *Bilingual Research Journal*, 45(2):180-204

Umansky I. M., Reardon S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms*. American Educational Research Journal*, 51(5), 879–912.

United Nations Human Rights Report 2021. Evaluation and lessons learned. Retrieved from: https://reliefweb.int/report/world/united-nations-human-rights-report-2021

U.S. Department of Education. Newcomer Tool Kit. (2016). Washington, DC. Retrieved from: http://www2.ed.gov/about/offices/list/oela/newcomers-toolkit/ncomertoolkit.pdf

US Government Accountability Office, 2020. Education in a Pandemic: The Disparate Impacts of COVID-19 on America's Students. Retrieved from: https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.pdf

Van Der Slik, F. W., Van Hout, R. W., & Schepens, J. J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PloS one*, *10*(11), e0142056.

Villegas, L., & Pompa, D. (2020). The Patchy Landscape of State English Learner Policies under ESSA. *Washington, DC: Migration Policy Institute*

Villegas L., Garcia A. (2021). A federal policy agenda for English learner education. *New America*. https://www.newamerica.org/education-policy/reports/a-federal-policy-agenda-for-english-learner-education/

Villegas, L. (2023). English Learner Funding Equity and Adequacy in K–12 Education. *New America*. https://www.newamerica.org/education-policy/briefs/english-learner-funding-equity-and-adequacy-in-k12-education/

Villegas, L. (2023). Accountability for long-term English learners. Phi Delta Kappan, 105(2), 13-18. https://doi-org.ezproxy.library.wisc.edu/10.1177/00317217231205935

WIDA English Language Proficiency Standards and Resource Guide: Prekindergarten-Grade 12. *Madison, WI: WIDA Consortium.*

WIDA Consortium. (2010). Assessing Comprehension and Communication in English State to State for English Language Learners: ACCESS for ELLs® Interpretive Guide for Score Reports. *Madison, WI: WIDA Consortium*.

WIDA Annual Technical Report for ACCESS for ELLs Online English Language Proficiency Test Series 503, 2021-2022 Administration https://wida.wisc.edu/sites/default/files/resource/ACCESS-Online-ATR-2021-22-redacted.pdf

Walqui, A. (2000). Contextual factors in second language acquisition. *ERIC Digest. ERIC Clearinghouse on Languages and Linguistics, Document ED444381*, Washington, DC

Warner, L. R. (2008). A best practices guide to intersectional approaches in psychological research. *Sex roles*, *59*(5), 454-463.

Weldon, S. L. (2006). The structure of intersectionality: A comparative politics of gender. *Politics & Gender*, *2*(2), 235-248.

Welton, A. D., Endo, R., & Hargrave, J. H. (2018). The Politics of Caring: Analyzing the Borders and Boundaries of Caring for Students of Color in PWI College Classrooms. *International Journal of Qualitative Studies in Education*, 31(2), 143-157

Zehler, A. M., Miyaoka, A., Chaney, B., Orellana, V., Vahey, P., Gibney, D. T., ... & Yilmazel-Sahin, Y. (2019). Supporting English Learners through Technology: What Districts and Teachers Say about Digital Learning Resources for English Learners. Volume II: Technical Appendices. *Office of Planning, Evaluation and Policy Development*, US Department of Education.

Zuberi T. (2001). Thicker than blood: How racial statistics lie. *University of Minnesota Press*.

Appendix A

Table A1. Descriptive Statistics: Overall Composite Scale Scores, ACCESS Online.

Grade	Statistic	2017	2018	2019	2020	2021	2022	2023
1	Average	282.51	281.81	279.41	276.60	270.26	264.17	264.57
	Std Dev	31.68	30.25	30.10	32.22	32.79	34.96	34.60
	N	161,884	173,593	181,012	174,515	141,732	181,047	183,803
2	Average	302.04	305.87	303.74	304.00	298.14	293.50	294.59
	Std Dev	32.33	31.19	31.88	34.26	33.45	34.42	34.47
	N	165,545	178,086	185,830	180,923	144,072	183,002	181,955
3	Average	320.20	323.27	321.64	322.20	315.86	311.60	311.21
	Std Dev	35.16	33.06	34.01	35.85	34.82	36.83	38.32
	N	173,951	184,300	185,973	178,447	143,678	182,746	177,660
4	Average	341.93	351.82	350.85	350.53	343.72	343.96	342.37
	Std Dev	33.02	32.21	32.36	35.09	34.87	37.54	38.48
	N	115,999	175,786	178,699	171,894	137,797	179,372	171,210
5	Average	347.26	354.17	357.40	355.83	348.83	351.11	346.68
	Std Dev	37.41	35.07	34.75	37.35	36.18	39.21	40.74
	N	84,876	114,285	142,882	137,687	107,205	154,494	140,440
6	Average	336.40	339.16	340.80	340.20	335.34	336.61	333.41
	Std Dev	35.76	33.60	33.80	31.80	29.74	31.75	32.45
	N	72,837	85,399	102,140	112,808	84,991	124,670	121,746
7	Average	344.10	344.86	345.01	344.19	343.46	340.96	339.59
	Std Dev	39.34	36.67	37.69	35.37	33.37	34.85	35.54
	N	74,444	81,202	91,017	101,836	85,456	117,750	122,276
8	Average	350.97	351.40	350.09	349.06	348.59	348.42	343.90
	Std Dev	42.13	38.86	40.46	38.59	36.29	37.93	38.02
	N	73,644	79,429	84,117	88,073	75,329	119,363	114,907
9	Average	355.04	358.72	357.74	352.23	359.09	350.83	350.56
	Std Dev	40.01	38.11	37.39	37.89	34.14	36.93	36.81
	N	92,048	90,105	96,070	99,133	64,788	126,354	131,558
10	Average	359.23	366.14	365.12	362.47	361.02	359.99	356.00
	Std Dev	35.75	37.86	35.68	35.48	35.47	35.22	36.24
	N	61,764	82,934	79,431	78,015	57,561	86,507	106,638
11	Average	364.11	370.27	370.94	369.16	367.60	363.95	362.46
	Std Dev	34.58	36.33	35.31	33.95	33.80	35.29	35.66
	N	43,395	59,322	70,319	64,028	46,166	72,648	75,763
12	Average	367.90	370.75	371.28	370.44	369.99	366.52	363.15
	Std Dev	33.45	35.49	34.72	33.34	32.76	34.08	34.99
	N	29,257	40,471	50,506	55,073	34,060	58,690	61,474

Table A2. Parameter estimates: Intersectional models (Ethnicity X Race).

Dependent Variable = CSS	Model 2: OLS		Model 3: XT/GLS		Model 7: MIXED	
Independent Variables	β	Robust SE	β	SE	β	SE
COVID-19	-7.73***	0.02	-5.32***	0.02	-6.63***	0.02
Years EL	9.38***	0.03	10.72***	0.02	9.91***	0.02
Years EL ^2	-0.57***	0.00	-0.63***	0.00	-0.62***	0.00
Newcomer	-1.96***	0.08	-7.36***	0.04	-6.71***	0.04
LTEL	-8.26***	0.06	-4.22***	0.05	-4.29***	0.05
SLIFE	-6.01***	0.02	-5.24***	0.02	-3.91***	0.02
Female	5.14***	0.04	6.55***	0.05	6.08***	0.05
IEP (Disability)	-21.47***	0.07	-12.18***	0.07	-16.03***	0.07
Migrant	-11.52***	0.32	-4.91***	0.24	-3.55***	0.24
LIEP Waiver	10.99***	0.18	5.40***	0.18	4.99***	0.18
Ethno-Racial Categories	*baseline category is 'No Race, not Hispanic'*					
Asian not Hispanic	19.91***	0.08	20.43***	0.06	14.87***	0.07
Asian Hispanic	8.21***	0.31	6.36***	0.34	6.53***	0.34
Black/African nH	8.33***	0.08	7.45***	0.07	8.22***	0.08
Black/African Hispanic	3.81***	0.17	1.43***	0.18	2.53***	0.18
Mixed/Multiple Races nH	17.31***	0.19	13.73***	0.19	10.86***	0.18
Mixed/Multiple Races Hispanic	2.52***	0.12	*-0.01*	0.12	1.24***	0.12
Native American or Alaskan nH	4.89***	0.13	0.59*	0.14	2.36***	0.16
Native American or Alaskan Hispanic	2.67***	0.09	-1.49***	0.08	0.88**	0.08
Pacific Islander or Nat HI nH	2.46***	0.12	1.37***	0.16	1.84**	0.16
Pacific Islander or Nat HI Hispanic	2.85***	0.15	-0.57*	0.17	0.38*	0.18
White nH	13.95***	0.08	10.82***	0.06	8.06***	0.06
White Hispanic	2.87***	0.08	0.32***	0.06	2.06***	0.07
Hispanic (no Race)	2.63***	0.08	-0.33***	0.06	2.24***	0.07
Ethnicity Interactions						
Hispanic Newcomer	-10.78***	0.09	-3.98***	0.05	-3.65***	0.05
Hispanic LTEL	6.05***	0.05	0.88**	0.05	2.19***	0.05
Hispanic Female	*0.06*	0.05	*-0.07*	0.06	-0.25*	0.05
Hispanic IEP	2.47***	0.08	2.00***	0.08	1.81***	0.08
Hispanic Migrant	8.31***	0.35	3.45***	0.26	0.86**	0.26
Hispanic Waiver	-2.05***	0.24	-2.25	0.22	-1.17***	0.23
Constant	260.34***	0.09	257.17***	0.06	260.34***	0.07
Grade fixed-effects	*baseline category is Grade 1*					
Grade 2	20.00***	0.04	19.12***	0.03	20.17***	0.03
Grade 3	33.24***	0.05	32.55***	0.03	34.64***	0.03
Grade 4	59.51***	0.06	59.95***	0.04	63.04***	0.04
Garde 5	66.07***	0.06	70.79***	0.04	74.79***	0.04
Grade 6	51.92***	0.06	58.84***	0.04	63.99***	0.06
Grade 7	57.43***	0.06	63.00***	0.05	70.12***	0.06
Grade 8	63.41***	0.06	69.07***	0.05	77.70***	0.07
Grade 9	72.85***	0.06	79.81***	0.05	88.36***	0.10
Grade10	78.14***	0.06	83.67***	0.05	93.77***	0.10
Grade 11	83.13***	0.07	87.24***	0.06	98.45***	0.11
Grade 12	83.07***	0.07	85.81***	0.06	98.22***	0.11
Random Effects / Variance						

State	-	-	-	-	29.41	7.86
District	-	-	-	-	81.96	2.47
School	-	-	-	-	113.37	1.15
Student	-	-	696.43	-	653.16	0.91
Residual	1034.27	-	353.44	-	375.47	0.68
N (observations)	9,683,892	-	9,683,892	-	9,683,892	-
n (students)	-	-	3,391,969	-	3,391,969	-
R-squared	0.51	-	0.7;0.4;0.5	-	-	-
ρ (AR1)	-	-	0.29	-	0.38	-

* p < 0.05; ** p < 0.01; *** p < 0.001

Table A3. Parameter estimates: auxiliary models ‘E’ (Ethnicity + Race).

Dependent Variable = CSS	**Model 2e: OLS**		**Model 3e: XT/GLS**		**Model 7e: MIXED**	
Independent Variables	β	Robust SE	β	SE	β	SE
COVID-19	-7.80***	0.02	-5.37***	0.02	-6.69***	0.02
Years EL	9.44***	0.03	10.74***	0.02	9.93***	0.02
Years EL ^2	-0.58***	0.00	-0.63***	0.00	-0.63***	0.00
Newcomer	-3.18***	0.08	-7.92***	0.04	-7.16***	0.04
LTEL	-8.84***	0.06	-4.42***	0.05	-4.34***	0.05
SLIFE	-6.18***	0.02	-5.36***	0.02	-3.92***	0.02
Female	5.10***	0.04	6.55***	0.05	6.08***	0.02
IEP (Disability)	-20.96***	0.07	-11.51***	0.07	-15.62***	-0.06
Migrant	-13.60***	0.32	-5.13***	0.24	-3.51***	0.10
LIEP Waiver	11.87***	0.18	5.75***	0.18	5.31***	0.17
Ethnicity and Race Categories	*baseline category is ‘not Hispanic’ for Ethnicity, ‘no Race’ for Race*					
Asian	11.45***	0.05	16.08***	0.06	11.28***	0.05
Black/African	0.13***	0.05	3.54***	0.06	4.75***	0.06
Mixed/Multiple Races	2.61***	0.08	3.63***	0.09	2.77***	0.09
Native American or Alaskan	0.44***	0.04	-0.30*	0.05	0.39***	0.06
Pacific Islander or Native Hawaiian	-3.15***	0.09	-0.61*	0.12	-0.49***	0.12
White	2.18***	0.03	2.98***	0.03	2.39***	0.03
Hispanic (any race)	-7.00***	0.04	-5.90***	0.05	-2.71***	0.04
Ethnicity Interactions						
Hispanic Newcomer	-9.41***	0.09	-3.44***	0.05	-3.20***	0.91
Hispanic LTEL	6.76***	0.05	1.12***	0.05	2.26***	0.33
Hispanic Female	*0.10*	0.05	*-0.06*	0.06	-0.25***	0.14
Hispanic IEP	2.08***	0.08	1.43***	0.08	1.40***	0.76
Hispanic Migrant	10.15***	0.35	3.59***	0.26	0.79***	1.24
Hispanic Waiver	-3.13***	0.24	-2.76***	0.23	-1.64***	0.51
Constant	269.04***	0.06	261.72***	0.06	262.37***	1.31
Grade fixed-effects	*baseline category is Grade 1*					
Grade 2	19.95***	0.04	19.09***	0.03	20.14***	0.03
Grade 3	33.16***	0.05	32.51***	0.03	34.59***	0.03
Grade 4	59.34***	0.06	59.87***	0.04	63.02***	0.04
Garde 5	65.88***	0.06	70.71***	0.04	74.72***	0.04
Grade 6	51.69***	0.06	58.75***	0.04	63.95***	0.06
Grade 7	57.22***	0.06	62.91***	0.05	70.00***	0.06

Grade 8	63.21***	0.06	68.99***	0.05	77.50***	0.06
Grade 9	72.61***	0.06	79.70***	0.05	88.17***	0.10
Grade10	78.00***	0.06	83.60***	0.05	93.50***	0.10
Grade 11	83.00***	0.07	87.19***	0.06	98.13***	0.10
Grade 12	83.04***	0.07	85.79***	0.06	97.94***	0.11
Random Effects / Variance						
State	-	-	-	-	33.08	9.12
District	-	-	-	-	82.92	2.50
School	-	-	-	-	115.48	1.16
Student	-	-	701.19	-	653.30	0.91
Residual	1038.13	-	353.44	-	375.95	0.68
N (observations)	9,683,892	-	9,683,892	-	9,683,892	-
n (students)	-	-	3,391,969	-	3,391,969	-
R-squared	0.50	-	0.7;0.4;0.5	-	-	-
ρ (AR1)	-	-	0.29	-	0.38	0.02

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table A4. Parameter estimates: main Intersectional models pre- and post-COVID-19.

Dependent Variable = CSS	Model 7: pre-COVID-19		Model 7: post-COVID-19		Post-Pre Impact
Independent Variables	β	SE	β	SE	β
Years EL	9.75***	0.03	10.92***	0.03	1.21
Years EL ^2	-0.64***	0.00	-0.69***	0.00	-0.05
Newcomer	-7.89***	0.06	-2.79***	0.07	5.10
LTEL	-4.45***	0.07	-5.44***	0.08	-0.87
SLIFE	-4.42***	0.03	-4.14***	0.03	0.29
Female	6.17***	0.06	5.64***	0.07	-0.48
IEP (Disability)	-18.36***	0.09	-17.61***	0.10	0.37
Migrant	-2.81***	0.34	-4.81***	0.36	-1.98
LIEP Waiver	5.87***	0.23	5.04***	0.29	-0.81
Ethno-Racial Categories	*baseline category is 'No Race, not Hispanic'*				
Asian nH	14.45***	0.09	16.72***	0.10	2.19
Asian Hispanic	6.85***	0.43	7.16***	0.49	0.23
Black/African nH	7.18***	0.10	10.91***	0.11	3.67
Black/African Hispanic	3.08***	0.24	3.30***	0.25	0.20
Mixed/Multiple Races nH	11.65***	0.25	12.63***	0.27	0.93
Mixed/Multiple Races Hispanic	2.27***	0.17	1.92***	0.17	-0.39
Native American or Alaskan nH	2.49***	0.20	3.99***	0.26	1.49
Native American or Alaskan Hispanic	1.46***	0.11	0.90***	0.13	-0.65
Pacific Islander or Nat HI nH	2.26***	0.20	2.85***	0.24	0.56
Pacific Islander or Nat HI Hispanic	1.38***	0.23	0.92***	0.28	-0.49
White nH	8.28***	0.09	10.39***	0.10	2.05
White Hispanic	3.01***	0.09	2.03***	0.11	-1.10
Hispanic (no Race)	2.95***	0.09	2.44***	0.11	-0.58
Ethnicity Interactions					
Hispanic Newcomer	-4.63***	0.07	-4.91***	0.08	-0.32
Hispanic LTEL	2.13***	0.07	4.50***	0.08	2.35
Hispanic Female	*-0.11*	0.07	-0.31***	0.08	*-0.19*
Hispanic IEP	1.84***	0.10	2.49***	0.12	0.69
Hispanic Migrant	0.96*	0.37	0.78*	0.40	-0.20
Hispanic Waiver	-1.54***	0.30	*-0.13*	0.37	1.39
Constant	263.51***	1.13	250.98***	0.87	-12.02
Grade fixed-effects	*baseline category is Grade 1*				
Grade 2	18.50***	0.04	21.92***	0.04	3.42
Grade 3	33.54***	0.05	34.80***	0.05	1.26
Grade 4	61.60***	0.05	63.22***	0.06	1.62
Garde 5	73.06***	0.06	74.00***	0.07	0.94
Grade 6	61.17***	0.08	62.06***	0.09	0.89
Grade 7	66.98***	0.08	67.36***	0.09	0.38
Grade 8	74.44***	0.09	74.28***	0.10	-0.16
Grade 9	83.84***	0.13	83.94***	0.13	0.1
Grade10	89.30***	0.13	88.86***	0.13	-0.44
Grade 11	94.10***	0.13	93.01***	0.14	-1.09
Grade 12	93.71***	0.14	92.29***	0.15	-1.42
Random Effects / Variance					

State	40.80	10.63	23.22	6.12	-17.64
District	79.61	2.49	83.95	2.56	3.95
School	104.01	1.14	94.29	1.08	-9.52
Student	582.45	1.45	640.99	1.81	58.59
Residual	391.56	-	353.07	-	-38.5
N (observations)	5,384,984	-	4,292,054	-	-
n (students)	2,336,453	-	2,239,549	-	-
ρ (AR1)	0.39	0.02	0.34	0.02	-0.05

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$