From Policy to Practice: Identification of State Approaches to Mesoamerican Language Identification

by

Karen Zuniga Zyskind

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Special Education

Dissertation Committee:

Dr. Beth Harn, Chair

Dr. Sylvia Thompson, Core Member

Dr. Lauren Cycyk, Core Member

Dr. Gabriela Pérez Báez, Institutional Representative

University of Oregon

Winter 2025

DISSERTATION ABSTRACT

Karen Zuniga Zyskind

Doctor of Philosophy in Special Education

Title: From Policy to Practice: Identification of State Approaches to Mesoamerican Language Identification

This study examined how states implement federal guidelines for identifying non-English languages, focusing on discrepancies and gaps in statewide language codes and processes used to document students' language backgrounds, with an emphasis on Mesoamerican languages. Using Critical Policy Analysis (CPA), the research analyzed how state policies address federal requirements for language identification and the documentation of Mesoamerican Indigenous languages. Iterative memo-writing techniques were employed to systematically categorize and analyze state policy data, uncovering patterns in the identification and recording of these languages. Additionally, a regression discontinuity model evaluated the impact of Oregon's updated language codes on the precise identification and documentation of Mesoamerican languages. Results indicate that states require more flexible student information systems, training, and resources to ensure accurate and comprehensive documentation of linguistic diversity, particularly for Mesoamerican languages. Limitations and future directions are discussed, along with recommendations for refining state-level language implementation of federal requirements.

CURRICULUM VITAE

NAME OF AUTHOR: Karen Zuniga Zyskind

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of California, Berkeley
University of Redlands
University of Oregon, Eugene

DEGREES AWARDED:

Doctor of Philosophy, Special Education, 2025, University of Oregon
Masters of Science, Communication Sciences & Disorders, 2018, University of Redlands
Bachelors of Art, History/Rhetoric, 2012, University of California, Berkeley

PROFESSIONAL EXPERIENCE:

Bilingual Speech Language Pathologist Consultant, Oregon Department of Education, 2023
Research Analyst, Center on Reinventing Public Education, Arizona State University, 2023
Bilingual Language Consultant, WestEd, 2022
Bilingual Speech Language Assessor, Riverside Unified School District, CA, 2018-2020
Site Based School Speech Language Pathologist, Riverside Unified School District, CA, 2018-2019
Reading Partners Site Coordinator, Garfield Elementary Redwood City, CA, 2014-2015
Cal Student BUILD Site Director, CAL Corps Public Service Center, 2011-2012

GRANTS, AWARDS, AND HONORS:

NIH Student Travel Award, Annual Symposium on Research in Child Language Disorders (SRCLD), June 2023

ASHA Research Mentoring-Pair Travel Award, Funded by grant #R13DC003383 from the National Institute on Deafness and Other Communication Disorders (NIDCD), July 2022

Davis-Bricker Award for Student Research, Funded through College of Education Scholarship at the University of Oregon, April 2022

UO Libraries Open Access Article Award, University of Oregon, 2024

Emerging Leadership Project, Winter 2022 Leadership Cohort, January 2022

Dynamic Measurement Group Award, Funded through the College of Education at the University of Oregon, March 2021

Write Now Fall Session Scholarship, National Center for Faculty Diversity and Development at University of Oregon, 2021

Project I-LEAD Recipient, Doctoral Training grant funded by U.S. Department of Special Education Office of Special Education Programs, September 2020 - Present

PUBLICATIONS:

Zyskind, K., Dorman, M., Medina, Y., & Báez, G. P. (2024). Visibility for Indigenous Students and Their Languages: Analysis of Home Language Data in Federal Reports across Seven U.S. States. Social Sciences, 13(8), 427. https://doi.org/10.3390/socsci13080427

Zyskind, K., & Macy, M. (2024). Authentic assessment for children with language considerations: early identification and connection to services. Perspectives on Early Childhood Psychology and Education, 8(1). https://doi.org/10.58948/2834-8257.1069

Hernández, Campos, I., & **Zyskind, K** (2023). Considerations in utilizing translanguaging practices to meet the language needs of maya children in U.S. Schools. Language, Speech & Hearing Services in Schools, 1–8. https://doi.org/10.1044/2022_LSHSS-22-00082

Durán, Wackerle-Hollman, A., Miranda, A., Chávez, C., Pentimonti, J., **Zyskind, K**., & Rodriguez, M. C. (2022). Spanish and English oral language growth rates of bilingual preschoolers: The Effect of Language of Instruction. Learning Disabilities Research and Practice, 37(3), 175–188. https://doi.org/10.1111/ldrp.12287

ACKNOWLEDGMENTS

This dissertation was only possible because of the tremendous support I received from my community, mentors, colleagues, and family. First and foremost, I want to express my deepest appreciation to my advisor, Dr. Beth Harn, for her unwavering guidance and patience throughout this journey. You were exactly who I needed to finish this program. Your encouragement and constructive feedback have been invaluable, especially as I navigated the complexities of policy and research. I also want to thank Dr. Lauren Cycyk, who has been an incredible mentor, guiding me through my first grant-writing experience and showing me what effective, compassionate research looks like. Dr. Sylvia Linan-Thompson, your insights into course design and instructional strategies have enriched my teaching abilities immensely—I am so grateful to have learned from you.To Dr. Gabriela Pérez Báez, thank you for always championing the linguistic rights of Mesoamerican children and for believing in my place in this space. Your passion and advocacy have inspired much of my work, and I am grateful for the chance to gain consultation experience under your mentorship.

Special thanks to my coders including Erick, Yessy, Maya, Cynthia, and Rafela, for your dedication and hard work. This study was possible because of you contributing your ideas, time and I would also like to thank the state representatives who took time out of their busy schedules to participate in my survey—your insights were instrumental in understanding the broader landscape of language identification.

I want to recognize my incredible friends, Alex, Meagan, and Tony, for standing by me through every high and low of this journey. You have all seen me at my bravest and at my most vulnerable, and your unwavering support means the world to me. Thank you for reminding me to take breaks, celebrate small wins, and keep moving forward.

To my family, thank you for being my foundation. To my Tía Chely and mom, Alma, who provided tireless childcare and endless love during the final stretch of my dissertation—I could not have done this without you. To my father, Pablo, who always encouraged me to pursue my passions, and to my sister, Daisy, thank you for reminding me that Latina mothers can balance it all. To my husband, Ari, thank you for your patience, love, and encouragement, and for being the steady anchor I needed. To my mother and father-in-law, Miriam and John, thank you for your unwavering support, guidance, and constant encouragement. Your reminders to pace myself and your help in preparing for conferences were invaluable, and I always knew I could count on you for anything I needed. To my Tía Debbie and Tío Nick, your weekly meal drop-offs were a constant reminder of the community that carried me through this journey, sustaining both my energy and determination to finish. And to my daughter, Carolina Paz, who joined me on the final leg of this journey—you are my greatest inspiration.

Lastly, I want to honor the Indigenous Mesoamerican communities whose linguistic diversity lies at the heart of this work. May we continue to uplift and learn from you about your languages and culture.

DEDICATION

I want to dedicate this to my Nana, my Tio Chuy, and all the first-generation scholars who learn from those who have passed before us. My Nana, who never learned to read or write, was the wisest woman I have ever known, and her strength and wisdom continue to inspire me every day. My Tio Chuy, who was the first in our family to become a lawyer, showed me the value of perseverance and dedication, and his legacy lives on in my work.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

The linguistic diversity of the U.S. student population is increasing, evidenced by the rising number of students from non-English-speaking households in U.S. public schools, which grew from 4.5 million students in fall of 2010 to 5.0 million students in fall of 2020 (National Center for Education Statistics, 2024). These children are referred to by many terms across academic disciplines, including multilingual learners, emergent bilinguals, bilingual learners, dual language learners, and English learners (ELs), the latter term being a federal designation used in U.S. public schools (Hammer et al., 2014). This dissertation focuses on multilingual children, or children exposed to more than one language regularly. These children often experience divided input across their languages, where their total language exposure is split among multiple languages. Consequently, they may consistently hear one language in specific contexts, like school or with specific communication partners, leading to varying proficiency levels in each language (Castilla-Earls et al., 2020; Hoff et al., 2012, 2018; Paradis, 2023). Given the variability in children's language exposure and its impact on their language and learning outcomes, the focus of this dissertation is on examining how states are set up to account for the linguistic diversity of students.

These children represent a diverse group, encompassing in language backgrounds, countries of origin, and whether they are US-born or immigrants. They also exhibit diversity in their racial or ethnic identities, educational experiences, and cultural backgrounds (National Academy of Sciences, Engineering, & Medicine, 2017). Often, this nuanced diversity is overlooked when using broad categorical terms like "English Learners" (ELs) Notably, past research has critiqued the use of "EL" terminology and classification, arguing that it adopts a

deficit-oriented perspective that focuses primarily on children's lack of English proficiency, often overlooking their linguistic and cultural assets (Flores et al., 2015; Gutiérrez & Orellana, 2006). Since the focus of this dissertation is on understanding the legal and policy frameworks in the K-12 educational space that impact them, the term English Learners (ELs) will be used throughout. It is done to adhere to the terminology used across federal and state educational agencies to ensure consistency in the analysis of the policy and practices in the United States (US). However, it remains imperative to recognize the rich linguistic and cultural diversity these children bring into their educational environments and to move beyond viewing these children as a monolithic group defined by the single characteristic of a child learning English.

The percentage of students identified as ELs in schools rose from approximately 4.5 million in fall 2011 to about 5.0 million by fall 2021 (National Center for Education Statistics, 2024). There has been persistent concern among educators and researchers regarding how to best support ELs due to clear disparities in their academic performance. Specifically, in academic subjects such as math, reading, and science, ELs typically obtain lower scores compared to non-EL students (Cook et al., 2011; Mancilla-Martinez & Lesaux, 2010, National Center for Education Statistics, 2024; Polat et al., 2016). ELs also confront significant systemic barriers in education, including reduced access to advanced and college preparatory courses, higher dropout rates, and lower college enrollment (Callahan, 2013; Estrada, 2014; Gándara et al., 2003). Given the educational hurdles these students face and their growing numbers in schools, there has been an intensified research focus aimed at ensuring ELs develop the language and academic skills they need to succeed in school and beyond. As a result, research on ELs has traditionally focused on children's acquisition of English proficiency and the length of time to their reclassification, or comparison of EL reclassification rates across different subgroups of ELs (Menken et al., 2012;

Thompson, 2015, 2015b; Umansky & Reardon, 2014). Other research has examined the impact of reclassification on students' academic achievements, their access to advanced coursework, and future career success (Slama, 2014; Thompson, 2015; Umansky & Reardon, 2014; Umansky, 2016). Additional research has probed the disproportionate representation of ELs in special education, exploring reasons why ELs are overrepresented or underrepresented within certain disability categories in special education (Artiles et al., 2010; Cruz & Rodl, 2018). Emerging research has begun to examine the intersections of children who have a disability and who are also categorized as ELs and their academic progress (Cooc, 2023; Kangas & Cook, 2020). Lastly, research has been conducted on ELs identification process with an emphasis on the use of a home language survey (Bailey & Kelly, 2012; Salerno & Andrei, 2021). However, limited attention has been paid to the accurate documentation of non-English languages, including the federal requirements and state practices (Zyskind et al., 2024). Accurate language documentation in schools is crucial to supporting ELs effectively, especially for children from Mesoamerican Indigenous backgrounds.

From 2000 to 2017, students designated as English learners (ELs) in U.S. public schools grew by over 30 percent (National Center for Education Statistics, 2023). Hispanic/Latino/e/x children represent 75% of ELs, with Spanish being documented as the primary language for 3.7 million of these students (National Center for Education Statistics, 2023). Importantly, there is evidence of underreported Mesoamerican Indigenous identities and languages within the Latino/e/x category. In the 2004 Census, 407,000 individuals self-identified as Hispanic American Indian—a number that surged by 344.7% to 766,112 people identifying as Latin American Indian by 2020 (Huizar, 2004; Pérez Báez et al., 2022). The 2020 Census introduced two new Indigenous categories, Aztec and Maya, which together accounted for 74.0% of the

Latin American Indian alone category (U.S. Census Bureau, 2023). Mesoamerican Indigenous languages and cultures include hundreds of languages and Indigenous groups beyond Maya and Aztec that are distinct from Spanish. Their presence in the US is not fully known, although there is increasing documentation in immigration courts and agriculture labor documents. However, their representation in schools is yet to be fully determined (Cambell, 1986; Obinna, 2023; Stephen, 2023; Goodrich et al., 2022). Consequently, Mesoamerican Indigenous children are often misidentified as Spanish English bilinguals, as they develop their English and Spanish skills in the context of a third Indigenous language – a misidentification and omission which has been calculated by one study to be done by a factor of 19 in schools (Martínez & Martinez, 2022; Martínez & Mesinas, 2019).

The accurate identification of a student's native language is a crucial step in the initial identification process for English Learners (ELs). Schools are tasked with identifying non-English languages to accurately identify potential ELs. However, it is unclear if current language documentation processes are accurate. Failure to correctly document a student's native language can lead to several issues. It can hinder the provision of appropriate language support services, as educational support and resources are often tailored based on the student's linguistic background (López, & Iribarren, 2014). Moreover, inaccurate language information may delay children's EL reclassification due to poor language support programming (Brooks, 2018; Menken et al., 2012). Additionally, incorrect language information can result in improper assessments of a student's language abilities and academic skills (Bedore et al., 2012; Grimm & Schulz, 2014). This may lead to inappropriate referrals for special education assessments that do not accurately reflect the student's entire linguistic repertoire, leading to compounding educational disadvantages.

Ultimately, failing to correctly identify a student's native language can exacerbate academic disparities for Indigenous student populations.

To address these academic disparities, it is crucial for more research to focus on understanding the linguistic profiles of children who use these lower-incidence Mesoamerican Indigenous languages. It is also important to understand the language documentation processes enacted in schools, as this can impact their academic trajectories significantly. To support ELs' language learning, educators need to comprehend children's cultural backgrounds, current language exposure, and consider the impact of language learning, variation, or disorder on their performance (Castilla-Earls et al., 2020; Oetting, 2018). Speech language pathologists (SLPs) and educators need to be aware of children's home language exposure and the exact languages children are exposed to (Guiberson & Atkins, 2012; Peña et al., 2021). Without this knowledge, educators and SLPs may express concerns about delays in children's acquisition of Spanish or English, unusual patterns in English productions that are not typical of Spanish-English bilinguals, and potentially contribute to the historical disproportionality of ELs in special education (Artiles et al., 2005, 2010; Goodrich et al., 2023; Martínez & Martinez, 2022).

Demographic Data Collected in Schools

State department agencies are required to gather demographic data from schools and local educational agencies, as mandated by Title III, Part A of the Every Student Succeeds Act (ESSA). The U.S. Department of Education (ED) uses two principal methods for this data collection: EDFacts and the Consolidated State Performance Report (CSPR). EDFacts centralizes data to inform education policy and decision-making, while CSPR monitors ESSA's implementation. To aid in fulfilling these requirements, ED provides a manual, the Guide to Collecting and Reporting Title III Data. This guide outlines the collection of essential

demographic data, including the number of immigrant students categorized by English Learner (EL) status, and detailed counts of ELs by grade level, native language, race/ethnicity, and disability status.

Demographic data, including enrollment, attendance, ethnicity, gender, and native language, are crucial for understanding trends and achievements within schools (Davis & Bauman, 2008; Matheny et al., 2023). Analyzing these data with other educational metrics allows researchers and educators to understand subgroup student performance. Analyzing demographic data supports schools in making informed decisions for educational improvements and interventions for student subgroups (Cuiccio & Husby-Slater, 2018; Young et al., 2018). Furthermore, demographic information plays a pivotal role in resource allocation, school zoning, and federal funding decisions (Castro et al., 2023; Renn, 2009). For example, these data help determine eligibility for federal programs like Title III, which provides language instruction to ELs and immigrant students; Title VII, which targets language support for traditionally underserved populations; and Title I Part C, known as the Migrant Education Program (MEP). Ensuring accuracy in demographic reporting is essential, as it impacts funding, program eligibility, and the support provided to meet the unique needs of students from diverse linguistic backgrounds. Clarity and guidance on state requirements for demographic data collection and reporting are imperative to maintain this accuracy.

Policy Context

The legal and policy framework supporting the education of ELs in the US has evolved over time, influenced by historical policies and legal precedents. The Civil Rights Act of 1964, the Equal Educational Opportunities Act of 1974 (EEOA), and the Bilingual Education Act of 1968 (BEA), along with landmark court decisions such as Lau v. Nichols (1974), Castañeda v.

Pickard (1981), and Plyler v. Doe (1982), have laid down a comprehensive policy landscape governing the federal requirements for the education of ELs. The 1964 Civil Rights Act, specifically Title VI, prohibits recipients of federal financial assistance, such as public schools, from discriminating based on race, color, or national origin. Title VI mandated that public school districts receiving federal funding provide educational programs for ELs that facilitate their access to academic instruction while learning English. The premise is that educational opportunities should not be denied based on language proficiency, nationality, country of origin, or immigration status. Consequently, it requires schools to communicate with families with limited English proficiency in a language they understand (Llahmon & Gupta, 2015; Miksch, 2019).

The next major federal policy impacting EL education was the passing of 1968 Bilingual Education Act (BEA, Title VII) which was the federal government's first recognition of their role to address challenges faced by EL students in schools (BEA; P.L. 90-247; Gándara, 2015). It provided funds for EL programs, planning implementation, training personnel and spurred state legislation supporting bilingual education (BEA; P.L. 90-247). The BEA was embedded within the 1965 Elementary and Secondary Education Act, reflecting a broader educational policy trend focusing on accountability and reporting by federal education programs. The 1974 reauthorization of the BEA clarified ambiguities in the original law. It emphasized equal educational opportunity and connected it with bilingual education programs, stating that the policy of the US is to ensure equal educational opportunities and to promote bilingual education practices. Bilingual education was defined as instruction in both English and, when necessary for effective progress, the student's native language. Later amendments to the BEA, such as the 1978 revision, emphasized English language acquisition, restricted the use of students' primary

languages in education, and set a two-year limit on participation in bilingual programs (Gándara et al., 2010).

The 1974 Supreme Court ruling in Lau v. Nichols set a significant precedent for the education of non-English-speaking students in the United States. The Court determined that providing equal access to facilities, textbooks, teachers, and curriculum does not constitute equal treatment under Title VI of the Civil Rights Act of 1964. The Supreme Court ruled that the San Francisco Unified School District had violated the Civil Rights Act of 1964, Title VI by not providing English language instruction to 2,856 Chinese-speaking children. The Lau Remedies, which were developed by the Office for Civil Rights to implement this decision, were used to apply to all other students whose first language was not English. This ruling emphasized that students who do not understand English are effectively denied a meaningful education unless specific actions are taken to accommodate their language needs. Following this decision, ELs were recognized as a protected class. The Supreme Court's ruling emphasized the importance for districts to proactively ensure effective education for English language learners aside from the standard resources available to all students.

The next landmark court cases were Castañeda v. Pickard (1981) and Plyer vs Doe (1982) which brought to the forefront the educational challenges faced by first and second-generation immigrant students. These students, who grew up speaking languages other than English, were previously placed in English-only classrooms without adequate support to learn the language or understand the curriculum, thereby missing out on educational opportunities. Castañeda v. Pickard established a three-pronged test to evaluate the effectiveness of bilingual education programs. The “three prong standard” included: (1) a program based on recognized theory; (2) faithfully implemented according to the theory, including adequate resources for

implementation; and (3) that demonstrated effectiveness over time. The Supreme Court, in the case of Plyler v. Doe, established that states are not allowed to withhold educational support from undocumented immigrant children. The court concluded that these children are entitled to receive a free public education from kindergarten to twelfth grade, overturning a Texas law that prohibited use of state funding in school districts to provide education to undocumented children. The ruling declared that the refusal to enroll and educate undocumented children infringed upon the Equal Protection Clause of the Fourteenth Amendment (457 U.S. 202, 1982).

The current educational legislative framework, outlined in Title III of the Elementary and Secondary Education Act (ESEA) of 1965—later reauthorized as the No Child Left Behind Act (NCLB) in 2002 and presently enforced as Every Student Succeeds Act (ESSA) of 2015—mandates that students identified as ELs participate in designated language assistance programs. Under ESSA, states are required to establish standardized criteria for identifying ELs students, implement appropriate programs, and monitor English proficiency measures through assessments, educational standards, and teacher quality. The current definition of English Learners stems from the ESSA reauthorization of 2015. Therefore, the current federal mandate is for all states to identify, evaluate, and monitor the English language skills of ELs until they achieve proficiency in English.

Prior EL Policy Frameworks

Raymaker Extant research examining education policies impacting ELs in public schools has underscored the importance of examining federal and state regulations and their implementation and potential impact on ELs. Policies examined include key areas such as identification, assessments of ELs, classification, and reclassification, a summary of which is provided next.

Lacelle-Peterson and Rivera (1994) critically examine assessment practices for ELs, emphasizing that education reforms must consider the unique demographic, linguistic, and educational needs of ELs. They challenge the assumption that reforms benefiting monolingual English students will automatically benefit ELs and instead call for the development of appropriate assessment policies and practices specifically tailored for ELs. This entails a collaborative approach to reform that involves EL specialists, educators, policymakers, and assessment experts in developing, monitoring, and revising assessment processes. They advocate for assessment practices that provide a comprehensive overview in both their native languages and English, emphasizing progress over time rather than relative scores at a single point.

Lacelle-Peterson and Rivera claim that an equitable education system for ELs must incorporate four fundamental elements: access, participation, benefit, and native language development (1994). First, it is crucial that ELs have access to the full range of content knowledge, ensuring they are not segregated into lower-track courses or denied the opportunity to engage with challenging subject matter. Second, meaningful participation in learning requires ELs to engage actively with challenging content, peers, and teachers, facilitated by instructional strategies that accommodate their language learning needs. Third, for ELs to benefit from and succeed in their education, schools must hold high expectations for these students, providing the necessary support and encouragement to help them succeed in advanced coursework and prepare for post-secondary opportunities. Lastly, the system should actively promote development of ELs' native language skills alongside their English development and consequently acknowledge the cognitive, cultural and societal benefits of bilingualism. Lacelle-Peterson and Rivera suggest that by prioritizing these elements, an education system can guarantee equity and excellence for

ELs, thus fostering their academic success and personal development within a diverse and multilingual framework.

Wiley and Wright (2004) analyze the historical impact of language policies on ELs, focusing on language and literacy policies as tools of social control. They critique resistance to bilingual education in the US, highlighting historical periods of tolerance of and resistance to bilingualism. They argue that historically, controlling people's language use was done to maintain social control, a practice evident as early as colonial times with African slaves and Native American peoples. Federal policies toward Native Americans included coercive assimilation measures like English-only boarding schools. The rise of the English-only movement affected bilingual education, which was common in areas with non-English speaking immigrant populations until World War I (Kloss, 1998). The federal government initially neglected bilingual programs but later promoted them with the Bilingual Education Act of 1968, leading to adoption across 30 states by 1990. Changes in 1974 clarified the role of native language instruction in facilitating academic progress alongside English language learning. The 1978 reauthorization introduced the term "limited English proficient" and lifted the ban on dual-immersion programs, yet a shift towards transitional bilingual education emerged. The 1984 reauthorization clarified Title VII program goals to help LEP students achieve English proficiency, introducing funding for English-only programs. In 1994, the Improving America's Schools Act increased attention to maintenance of bilingual programs by creating new grant categories to support bilingual education such as personnel training grants, professional development for educators, and grant funding to promote bilingual program development and implementation. The No Child Left Behind Act of 2002 replaced Title VII with Title III, emphasizing English proficiency without explicit support for maintaining bilingual programs.

This shift on bilingualism decreased funding per eligible student and incentivized districts to transition to English-only instruction, despite established bilingual education benefits. By tracing laws and policies, Wiley and Wright urge that researchers and practitioners reevaluate language policies to recognize diversity and promote bilingual education for better serving language-minority students.

Linquati and Cook (2013) propose an organizing framework for a unified EL definition, emphasizing the need for consistent identification and classification procedures to optimize federal funding allocation for state EL instructional programs. Their four-stage framework includes identifying potential ELs, confirming EL status, setting English proficiency standards, and reclassifying ELs. By aligning EL definitions across consortia and states, they aim to improve the accuracy of initial classification procedures and propose evaluating the reliability and validity of each state's Home Language Surveys so that the focus is on understanding a child's current language use and degree of exposure to English. They also stress the implication of using federal EL definitions and civil rights case law to uphold the rights of students classified as ELs. For instance, they note a major limitation in the definitions of the Civil Rights Act and Equal Education Opportunities Act. While it establishes educational rights for language-minority students, these laws do not clearly define what constitutes a child who has limited proficiency in English nor how to assess it. According to Linquati and Cook, when looking for federal guidance, it is the Elementary and Secondary Education Act which provides the clearest definition, stating that students with limited English proficiency are those who are impacted by their non-English linguistic backgrounds to an extent that may hinder their academic achievement. Linquati and Cook hence advocate for creating a uniform definition of ELs across states and assessment consortia to ensure equitable resource or funding allocation. They also

acknowledge the challenges in setting performance standards for English proficiency assessments and propose the creation of standardized assessment tools to ensure consistent reclassification criteria. They advocate for further research to enhance decision-making processes in these areas. Overall, Linquati and Cook provide a framework for states and consortia that supports the establishment of a common EL definition, one that is clear, consistent and aligned with federal guidelines, centered on ensuring that children from non-English backgrounds have access to necessary resources and funding to succeed academically.

Bailey and Carroll (2015) present a framework structured around four key components: initial identification via the Home Language Survey, standardized screening, annual proficiency assessments, and reclassification procedures. Their framework aims to establish a cohesive system for effectively monitoring and supporting ELs language and academic progress. They stress the importance of adopting a systems-view to recognize the intricate connections among these components and promote collaborative efforts from federal and state agencies down to the school-level, involving educators and support staff to enhance ELs educational experiences. Critically, Bailey and Carroll underscore the need to refine initial identification processes and advocate for a nuanced approach to the allocation and monitoring of Title III services. They argue for the development of more effective instructional assessments beyond standardized testing to mitigate misclassification risks and bias. Additionally, they propose comprehensive validity studies and balanced reclassification standards to facilitate appropriate transitions for ELs out of language support programs. Finally, they call for assessment systems that are reflective, data-informed and aligned with children's language development and academic achievement goals.

Pompa and Villegas (2017) were the first to create a framework for states to use to develop, structure and evaluate their state plans of EL accountability for K-12 academic achievement. Their framework stresses the creation of uniform long-term goals across student subgroups, inclusion of ELs in academic achievement goals, with additional indicators for elementary and middle school graduation rates for high schools. Pompa and Villegas emphasize that ELs are heterogeneous, possessing diverse starting points and experiences before entering K-12 education. Due to this diversity of experiences, they underscore the importance of state plans to adequately account for students with unique experiences, such as those with limited or interrupted formal education, long-term ELs and newcomer students. Moreover, the framework addresses the need for states to develop interim progress measurements, stressing states create assessment options for newly arrived ELs. The core of the framework stem is to emphasize transparency and inclusivity in evaluating school performance such that it accounts for all ELs, advocating for evidence-based interventions tailored to support subgroup needs. Overall, this framework provides what it means for state leaders to analyze state plans for ELs, ensuring effective support and accountability for EL academic achievement.

Umansky and Porter (2020) propose a framework aimed at guiding state-level policies, prioritizing equity and recognizing linguistic diversity as an asset, particularly focusing on defining the state's role in implementation of federal policies. According to Umansky and Porter, the state is responsible for establishing basic protocols for EL programming by setting standards and certification requirements, funding accreditation, and overseeing and specifying education programs in relation to EL education. Their framework is organized around three state-level principles for EL education: understanding student needs and assets, providing accessible high-quality instruction, and supporting effective and aligned K-12 education systems. Understanding

student needs and assets involves recognizing and responding to the diverse backgrounds, skills, and needs of ELs. This relates to EL diversity, acknowledging the wide range of linguistic, cultural, and educational backgrounds among ELs. It also includes policies related to assessment practices that accurately reflect EL students' knowledge and skills, ensuring that assessments are valid and reliable for EL students across content and language domains. Additionally, it involves policies concerning classification and reclassification procedures that fairly identify ELs and guide their transition out of EL status when they are ready, based on clear, equitable criteria. Moreover, once identified, the framework stresses the importance of providing ELs with access to high-quality instruction, ensuring they are not excluded from challenging coursework and have they receive the necessary support to succeed in those classes. This support involves integrating English language development instruction alongside content learning, equipping EL students with the English language skills they need to engage with the content matter. Notably, this support can also include bilingual education programs, where English language development and home language maintenance occur simultaneously as children access core content. The final principle is supporting effective and aligned education systems stresses the need for education systems to work together cohesively to support EL education. Key policy areas include teacher preparation, ensuring educators are trained and equipped with the knowledge and strategies to support ELs effectively. It also emphasizes the need for responsive funding systems that provide the resources necessary for high-quality EL education and for alignment between pre-K through postsecondary education systems, ensuring smooth transitions for EL students across educational levels and maximizing their educational outcomes.

As the previous policy frameworks demonstrate, language and education policies hinge on the accurate identification of potential ELs. This initial identification is crucial, not only

because it triggers a series of actions to meet federal support requirements but also because it shapes the entire educational experience of children learning multiple languages. Importantly, each framework calls for a systematic and coordinated process to support ELs, spotlighting potential gaps where the current system may fail to meet their needs adequately and the complexity of these students' needs. Unfortunately, aside from the use of the Home Language Survey to identify non-English languages, the conceptual frameworks fail to examine how these languages are identified and documented in schools, which is pivotal for refining the subsequent EL identification and monitoring mechanisms. The conversation typically centers around the methods used to identify potential ELs, underscoring the need for more comprehensive research and policy development in this area.

Language Identification

Under the ESSA, states must identify students with a native language other than English or from environments where a non-English language may have influenced their English language development. The identification of ELs primarily hinges on two types of information: the Home Language Survey (HLS), and the English proficiency assessment. While all 50 states use the HLS for this purpose, it is important to note that the contents of these surveys are not federally required (Baily & Linquanti, 2014). Yet, the methodologies for identifying and classifying ELs exhibit considerable variation across different states and school districts, leading to inconsistencies in their educational experiences (Abedi, 2008; Linquanti & Cook, 2015). This initial classification not only dictates the student's immediate academic path, such as eligibility for language support programs, but also labels them as an EL, impacting their educational journey (Umansky, 2016, 2017; Umansky et al., 2017). ELs are identified upon school registration if they use a language other than English at home, leading to an assessment of their

English proficiency across reading, writing, speaking, and listening, irrespective of their immigration status or nationality (Abedi, 2008). Students failing to meet the proficiency benchmarks are classified as ELs and are entitled to additional support to enhance their English skills across all domains, thereby ensuring they can meaningfully access and participate in educational programs (ESSA, 2015; U.S. Department of Education, 2016). Those who exceed the thresholds are deemed to not require additional English support. Importantly, after testing, parents or caregivers are allowed to decline or waive English language support services after being notified of their child not passing the English proficiency threshold. The notification will include information about available language support programs, the option to waive English language support services, the right to withdraw the child from English language services, and the criteria for exiting English language services (Office of Elementary & Secondary Education, n.d.). Notably, unlike other federal policies such as special education law (Individuals with Disabilities Education Act [IDEA]), this policy operates on an opt-out-model, where parents or legal caregivers can opt out of EL services, but they are not required to actively agree to them, nor are there formal meetings or opportunities for dialogue between families and schools to make decisions about EL services (Thompson & Rodriguez-Mojica, 2022).

Previous research on the use of HLS highlights their inadequacies in capturing complex language learning experiences of children before they begin formal schooling (Bailey & Kelly, 2012; Salerno & Andrei, 2021). While ESSA mandates states to identify if a child is exposed to a non-English language, there is no strict guidance to how this identification takes place, leading to a wide range of HLS questions used across states (Bailey & Kelly, 2012; Durán, 2008). Importantly, the HLS may not properly account for the diverse and complex backgrounds of EL students. ELs may vary significantly in their sociocultural contexts, proficiency levels in their

multiple languages, preference for language use, educational histories, and the quality of previous language instruction received (National Academy of Science, 2017). Additionally, the HLS may pose challenges to language identification, especially for students who speak less recognized languages beyond major ones like Spanish or English.

Bailey and Kelly (2012) examined state procedures for HLS use and implementation guidance from all 50 states. They emphasize the importance of the collaboration between the Office for Civil Rights (OCR) and state education agencies to ensure compliance with laws. Through analysis of state procedures and case studies from six states, they identified concerns with HLS implementation, including issues with the order of language acquisition and ambiguity in survey items. They questioned the assumptions behind the HLS questions, emphasizing the necessity of validation studies for HLS. They pointed out that the HLS might not accurately capture the diversity within the student population, potentially overlooking linguistic minorities and balanced bilingual speakers. They found that current surveys are primarily designed to measure whether a language other than English is used in the home and proposed refining HLS forms to concentrate on current language proficiency and language exposure, which they believe will lead to better accuracy of language exposure for ELs. They conclude by emphasizing the need for comprehensive analysis of state practices and the provision of concrete advice to improve HLS implementation and interpretation.

Salerno and Andrei (2021) conducted a review of HLS across all 50 states, focusing on data from spring 2020. They collected HLSs from state education department websites, using WIDA versions when not available online and contacted the state agencies directly when needed. Salerno and Andrei organized the HLS questions into a spreadsheet, coding them to assess how states identify students for language screening, and excluded non-screening questions. States

were categorized based on HLS question alignment with US Department of Education recommended guidelines for inquiring about language use, which is typically done by asking: "what is the language most often spoken by the student?" Results showed variability in HLS questions and interpretations, which would affect the identification process for EL students. While most states require further screening for non-English languages, variations exist. To examine the impact of this variation, Salerno and Andrei examined six fictitious case students, uncovering inconsistencies in HLS identification, particularly for those with complex language backgrounds. For example, fictitious students, Eva and Valentin, who have a clear primary language background of Spanish were consistently identified by all HLS as having Spanish as their first only dominant language. In contrast, children with more nuanced language exposure, including bilingual children or children with varied exposure to more than one language, showcased the HLS limitations. For instance, in the case study of Stefan, a fluent bilingual with consistent exposure to English and Bulgarian would only be identified by 27 states for further screening and the HLS would fail to capture both languages used in Stefan's home. Similarly, the case study involving a heritage Spanish learner, Raul, who primarily uses English in the home led to varied results with 17 states identifying him as needing further testing, and 27 others provided unclear guidelines due to varied family language practices in Spanish. Lastly, Peter, a foreign exchange student, would be identified by 48 states as EL. In all, Salerno and Andrei underscore the importance of consistent HLS questions across states to address inconsistencies for students with complex language backgrounds so proper supports can be deployed and to accurately understand the language complexities in our schools.

Problems with Language Recording

According to ESSA, section F among the State Plan Requirement, each state's education plan (referred to as "State plan") must include specific provisions related to language assessments. One main requirement is the Identification of Languages: “Each state plan shall identify the languages other than English that are present to a significant extent in the participating student population of the state” (ESSA, § 1111(b)(2)(F)(i)). Each state must thus document and report on the linguistic diversity within its student body to the U.S. Department of Education (ED). The ED provides guidance on language reporting requirements through a Guide titled Guide to Collecting and Reporting Title III Data (Guide) (Office of Elementary and Secondary Education, 2024). These requirements lay out that States must use ISO 639-2 Code List for language classification. The ISO 639-2 Code List is a system designed to standardize language representation across diverse educational settings, and which is overseen by the United States Library of Congress. Language codes were developed by the International Organization for Standardization (ISO) established to uniquely identify languages across databases, as an attempt to standardize and simplify language representations across different spellings and naming convention (DeCamp, 2008; Simons, 2013). ISO 639-1 was established in 1988 and updated in 2004. It introduced a two-character language code. ISO 639-2 was introduced in 1998 and expanded the naming convention of language to the use of three-letter codes (Library of Congress, 2017.). In 2001, SIL International began to develop the ISO 639-3 with the aim to align the codes with Ethnologue, the Languages of the World catalog which details language use by region, country, and identifiers of dialects and sub-language variations. Since completion in 2007, the ISO 639-3 has been deemed the standard coding system to document languages and is maintained by a global collaborative of linguists, allowing for the creation and modification of new codes (Drude, 2018; SIL International, 2024). ISO 639-3, like ISO 639-2 uses three-letter-

code. However, unlike ISO 639-2, 639-3 focuses on individual languages and covers a larger number of lesser-known languages compared to the ISO 639-2. While this list is available and in use across contexts, the current framework available for SEAs is the more limited ISO 639-2. Currently, the US Department of Education denotes ISO 639-2 as the standard language coding system to use to report language information for EDFacts data submission.

The ISO 639-2 Code List exhibits several limitations in its approach to categorizing languages. Notably, it tends to group language families, branches, and individual languages indiscriminately, leading to inconsistencies in the level of granularity. Despite there being an estimated 8,000 languages worldwide (Hammarström et al., 2023), the ISO 639-2 list only offers 500 three-letter codes to label languages, which limits and leads to omission of many less recognized languages. Furthermore, the list demonstrates a bias towards prominent global languages and their families. For instance, it provides specific codes for widely spoken languages such as Spanish ('spa') and French ('fre/fra'), and even a broader category code for Romance Languages ('roa'), which encompasses both Spanish and French. These languages are further categorized under the larger family of Indo-European Languages ('ine'), which also receives its own code (Zyskind et al., 2024). This level of granularity is not extended to all language families. For instance, the ISO 639-2 includes only five codes that can be used to categorize Mesoamerican languages: the Mayan language family, the Zapotec branch of Otomanguean languages, the Nahuatl branch of Uto-Aztecan languages, and Central American Indian languages. No individual languages within these different groups have a dedicated, individual ISO 639-2 code.

As previously stated, this study focuses on Mesoamerican languages and understanding the current approach to language documentation in schools from the federal requirements to state

level practices. It is important that Mesoamerican languages are documented with greater granularity to ensure languages are not rendered invisible behind broad labels or broader language categories (Stephen, 2023). When languages are categorized too broadly, it can lead to a lack of recognition and support for specific linguistic communities. This invisibility can result in inadequate educational resources, insufficiently tailored instructional strategies, and a failure to address the unique cultural and linguistic needs of students. What follows is a brief overview of the linguistic diversity of the Mesoamerican region.

Mesoamerican Context/Diversity

The Mesoamerican region, stretching from central Mexico to northern Central America (encompassing Guatemala, Belize, Honduras, El Salvador, and Nicaragua), forms a Sprachbund (Campbell et al., 1986) that is linguistically and culturally distinct from Latin American language and culture (Campbell, 1983; 1986; López-Austin & López-Luján, 2001). Mesoamerican Indigenous languages represent vast and grammatically distinct linguistic systems that vary greatly from Spanish and English. Mesoamerican languages share some linguistic traits, such as marking possession directly on a noun, use of a relational noun such as a body part term to fulfill syntactic functions otherwise fulfilled by prepositions in other languages, and use of verb subject object word order (Campbell et al., 1986; Suarez, 1983). These languages encompass several hundred varieties from 11 major language groups (Campbell, 2017; Suarez, 1983). These range from isolates like Purepecha, Huave, Xinca and Oaxaca Chontal, each of which is not related to any other language family, to complex language branches such as Uto-Aztecan language family consisting of approximately 31 Nahuatl languages, or the Mayan language family consisting of over 30 distinct Mayan languages, Totonacan-Tepehua consisting of 10 distinct languages, and the Mixe-Zoquen language family with 19 distinct languages (Campbell & Kaufman, 1985;

Campbell et al., 1986). Additionally, there is also the extensive array of Otomanguean languages, totaling over 180 distinct languages (Suárez, 1983). It's important to note that Mesoamerican Indigenous languages have no linguistic connection to Spanish. In fact, the diverse communities speaking these languages predate the arrival of the Spanish language brought by European colonization by thousands of years (Campbell et al., 1986, Suárez, 1983)

The diversity within Indigenous Mesoamerican languages cannot be overstated (Campbell, 2017; Pérez Báez, 2013; Suárez, 1983). This complexity can be seen with the Otomanguean language family, as depicted in Figure 1 by Campbell (2017) as adapted from Kaufman (1988). The Otomanguean language family divides into the Western and Eastern Branches which further splits into four additional subbranches where upon these four subgroups bifurcate into eight major subgroups that are comparable to Indo-European subgroups like Romance or Germanic (Campbell, 2017). The language branches seen at the end of the figure do not indicate the end of the branch but expanded further as seen in Figure 1.

To illustrate the importance of adequate language labels and coding, two language branches in the Otomanguean language family will be addressed. There is immense linguistic complexity exemplified by the Mixtec branch, which includes 53 languages (Hammarström et al., 2024). See Figure 2. Often mistakenly called a single language, 'Mixtec' comprises a complex array of languages with a history of diversification spanning over 2,000 years. These languages have developed distinct linguistic features that render them largely unintelligible to one another (Josserand, 1983). Use of a broad label like “Mixtec” would be like using the label Romance in place of the labels of Italian or Portuguese. Similarly, the Zapotec branch is not a single language but a branch of languages that, while sharing some common features, are generally not mutually intelligible. See Figure 3. Each Zapotec language possesses unique linguistic characteristics

specific to the community where it is spoken (Pérez Báez, 2013). Use of a broad label like Zapotec likewise would result in rendering the exact language used invisible. Therefore, it is essential that Zapotec languages be documented with the finest level of detail as using the broad term Zapotec would obscure over 56 distinct languages (Campbell, 2017).

Diaspora and Immigration

The history of immigration and diaspora among Indigenous peoples from the Mesoamerican region to the US is gaining increased scholarly attention, extending beyond the traditional Latinx context (Fox et al., 2004; Velasco Ortiz, 2016). Researchers suggest examining Indigenous mobility separately from that of Latinx populations, delving into the ethnic diversity within these Indigenous groups. This focus prompts inquiries into Indigenous displacement, mobility inequalities, and the multidimensional aspects of Indigenous lives, including their cultural practices, languages, and beliefs (Pulido et al., 2012; Velasco Ortiz, 2016). A crucial aspect of understanding the current landscape of Mesoamerican languages in schools is recognizing the longstanding presence of Mesoamerican Indigenous peoples in the United States, demonstrating that their languages and customs are not new to the US, but now may be better appreciated.

Historically, the presence of Mesoamerican Indigenous people in the United States dates back to the mid 1800s at the conclusion of the US-Mexico war, which resulted in Mexico ceding a large portion of its territory to the US (Stephens, 2023). This period marked the beginning of interconnected growth in US and Mexican industries such as mining, ranching, and agriculture, which created opportunities for labor migration (Acuña, 2007; Stephens, 2007). The Bracero Program, an agreement between Mexico and the United States between 1942 to 1964, allowed Mexican nationals, including Indigenous peoples, to work in the US on short-term agricultural

labor contracts (Cohen, 2011; Loza, 2016; Stephens, 2023). Additionally, the Immigration Reform and Control Act of 1986, and the Special Agricultural Workers Program, facilitated legal residency for many farmworkers, including Indigenous people (Stephens, 2023). The end of the twentieth century was marked by further displacement in the Mesoamerican region to the US due to a period known as "la violencia" in Guatemala (1970s-1991) which was part of the broader civil war lasting from 1960 to 1996 which led to many Indigenous people to seek refuge from violence, and which led to the displacement of 1.5 million people (Sanford, 2003). The 1990s' also witnessed an increase in Mexican immigration to the US, driven by significant political and economic changes, primarily related to the passage of the North American Free Trade Agreement (NAFTA), which prompted changes to Mexico's property laws in 1992 which lead to the dissolution of the ejido system, communally used lands, and privatized them taking away their ability to farm. These developments prompted migration to the US, especially from rural and Indigenous communities (Dunn, 2000; Fernández-Kelly et al., 2007; Holmes, 2013; Massey & Sana, 2003).

Transitioning to modern immigration trends, recent data from the National Center for Education Statistics (NCES) shows a rise of EL students in newer destination states, particularly in the South, such as Louisiana, Mississippi, West Virginia, and South Carolina (NCES, 2022). Between 2009-2015, EL student populations grew by 42.7% to 83.5% in the first three states, while South Carolina saw an increase of 83 percent. Furthermore, migration from Central America, specifically from El Salvador, Guatemala, and Honduras, has risen over time, with a large portion consisting of undocumented immigrants, including unaccompanied minors (Krogstad et al., 2019; U.S. Department of Homeland Security, 2020). Moreover, there has been a rise in the number of immigrants emigrating from Latin America who identify as Indigenous;

however, their experiences and languages often remain unrecognized by US institutions (Riva, 2022). Notably, about half of the 250,000 immigrants from Guatemala apprehended at the US-Mexico border in 2019 were of Mayan descent and did not speak English or Spanish (Ahtone, 2018; Nolan, 2019). In fact, Mam, K'iche', and Q'anjob'al—Mesoamerican Mayan indigenous languages—are among the top 25 most common languages used in immigration courts in recent years (Medina, 2019; TRAC Immigration, 2021). Yet, information about these indigenous languages in public schools remains scarce (Campbell-Montalvo, 2021,2023). Notably, Mayan languages such as Mam, K'iche' and Q'anjob'al are not currently listed as individual languages in the ISO 639-2, instead the broad label for Mayan language family is available for use by SEAs (LOC, 2017). Additionally, the National Center for Education Statistics (2024) reports that home language data is incomplete for Hispanic children who speak languages other than Spanish at home, indicating a gap in recognition for languages indigenous to Latin America.

Educational Research on Mesoamerican Children in US Schools

Velasco (2010) presents a case study that explores the educational experiences and challenges of two Indigenous Mexican children in a Spanish-English bilingual classroom in a New York City school. Velasco provides an overview of the demographic and linguistic diversity among Mexican immigrants in New York City and New Jersey. She notes the increasing presence of individuals from regions like Puebla and Oaxaca, where languages such as Nahuatl, Mixteco, and Zapotec are commonly spoken. She notes that Mexican Indigenous people in the US face unique challenges. These include issues with Census data collection due to language barriers. Additionally, there are educational misclassifications stemming from the incorrect assumption that these Indigenous immigrants predominantly speak Spanish at home. Velasco discussed two Mixtec children to illustrate the mismatch between the rich oral traditions

of their Indigenous language practices that would impact their educational experience. She discusses how their language which can involve the use of non-linear narratives and can include redundant structures, advice-giving and myths compared to the structured, temporal linear expectations of Spanish and English academic language. Ten observations from a dual-language kindergarten program were conducted to demonstrate the crucial role of teacher mediation in fostering academic language development. Ms. Stella, a teacher in the bilingual program, recognized the children's Indigenous background and tailored her approach to respect and integrate their linguistic and cultural heritage. She engaged in daily brief conversations with each child, using strategies such as asking precise questions and interactional scaffolding to help the children provide clearer narratives of his experiences, and recasting and stretching language to help both children learn new words and structures. Velasco ultimately contends that the effectiveness of bilingual programs for Mexican Indigenous children hinges on making pedagogical changes that successfully bridge the gap between their oral traditions and the literacy demands of school.

Bishop and Kelly (2013) conducted ethnographic research that explored the language-related decisions Indigenous Mexican migrants in New York City implement in their homes and with their families. The study centered on the experiences of three Indigenous Mexican mothers Marisol, Francisca, and Elena, who immigrated to New York City from Guerrero, Mexico. They spoke Spanish and an Indigenous language either Mixteco or Nahuatl. Through semi-structured interviews and language maps, participants traced their family's language history and current language experiences in the US. The authors reported that discrimination and belittlement faced by the participants as speakers of Indigenous languages influenced their language use decisions. For instance, one parent, Francisca, recounted instances of being ridiculed and being called an

"indio," a derogatory identity term after her use of Nahuatl in public. This discrimination led her to speak Spanish in public and to avoid teaching her children Nahuatl to protect them from similar experiences. Bishop and Kelly reveal that while the three Indigenous Mexican mothers reported being proud of their heritage languages, they prioritized teaching Spanish and English to their children due to perceptions of power and opportunity. Despite valuing multilingualism, each parent disclosed fear their children might not excel in multiple languages or speak Spanish or English "perfectly." Additionally, each parent expressed that migration to the US was aimed at improving their families' lives, but it came at the cost of discrimination, and linguistic and cultural loss.

Relatedly, Ruiz and Barajas (2012) conducted a qualitative study to explore the educational experiences of Mexican Indigenous students in Central California. Initiated by the Office of Migrant Education in California, the study addressed a notable gap in research concerning the educational needs and challenges faced by these students. Using qualitative methodology, including interviews and questionnaires, the researchers gathered data from diverse participants, such as elementary school principals, Migrant Education teachers, an Indigenous instructional aide, and a family from Oaxaca consisting of both adults and school-aged children. The study focused on three primary areas: identifying the strengths that Mexican Indigenous students and their families contribute to schools, examining the challenges they encounter within educational and community settings, and proposing recommendations to enhance their educational experiences. Key findings revealed Indigenous students experienced pervasive discrimination, both in their home country and in the US, specifically concerning Indigenous language use. This discrimination often manifested in instances of bullying and inhibited the students' expression of their linguistic and cultural identity at school. Additionally,

the study uncovered a divergence within the Indigenous community, with some individuals embracing assimilation into Mexican culture while others pursued integration into the dominant American culture and English language, viewing it as a pathway to improved opportunities. Additionally, Mexican Indigenous parents confirmed a preference for bilingual education, often because it allowed them to assist with homework using any Spanish skills they may have. Notably, administrators and teachers suggested that Indigenous children exhibited English learning strengths. In all, by engaging with school personnel, Indigenous families, and educators, Ruiz and Barajas illustrate that the Mexican Indigenous student population has multifaceted needs and point to future research areas for these students concerning Indigenous language use.

Machado-Casas (2012) conducted a three-year qualitative study aimed at understanding the experiences of Indigenous Latino immigrants in the United States, focusing on their identity adaptations across different social contexts. Drawing on the metaphor of a chameleon from the song "El Camaleón" by Los Altos de la Sierra, Machado-Casas illustrates how these immigrants adeptly shift identities for survival—physically, socially, and culturally. Data was collected through monthly meetings with Latino immigrant parents at six urban schools in North Carolina, involving 230 participants predominantly from Mexico, El Salvador, Guatemala, and Honduras. These participants represented various Indigenous communities, speaking 17 different languages and encompassing diverse Indigenous backgrounds including Otomi, Maya, Nahuatl, Zapotec, Mixtec, and Tzotzil. Key insights from the study were gathered from the experiences of three participants: Iza, an Otomí woman from Mexico, Maria, a Quiché woman from Guatemala, and Manuel, a Náhuatl/Pipil man from El Salvador. The author identified two main survival strategies these three participants reported using for survival: "coyote tools," involving the adoption of disguises or the use of multiple languages and cultural identities to blend in, and

"cruzando fronteras" (crossing borders), which refers to the use of fluid identities to maintain connections with their communities of origin. She states that despite the linguistic diversity and unique cultural backgrounds represented, schools and programs largely lacked awareness of these details, resulting in a lack of initiatives tailored to the linguistic and cultural needs of these children. Furthermore, many parents intentionally avoided standing out, contributing to their invisibility within these contexts. Machado-Casas concludes that Indigenous Latinos remain largely unseen and emphasized the need for further research and efforts from both researchers and educators to understand the adaptability and resilience of these populations.

Baquedano-Lopez (2019) conducted an ethnographic study to explore the migration experience of Indigenous Maya families with children from Yúcatan, Mexico, focusing specifically on their integration into a school in the San Francisco area. Baquedano-López discusses the process of "Latinization" experienced by Indigenous Mayan children in school, where they are often categorized as Latino and Spanish-speaking, overlooking their actual language and cultural backgrounds. Baquedano-Lopez' study is centered on understanding the migration patterns of indigenous immigrants from Yucatán and was conducted in two phases. The first phase, conducted from 2008 to 2009, involved fieldwork in Yucatán with return migrant families and various societal stakeholders. The second phase, from 2013 to 2016, focused on the educational and linguistic experiences of indigenous Yucatec students at Metropolitan Elementary School and its afterschool program. Baquedano-Lopez also explored how school administrators and teachers responded to the presence of Indigenous Mayan students, their inclusion strategies, and their perceptions of indigeneity. In their findings, Baquedano-Lopez notes that the San Francisco school they collaborated with does not specifically record Indigenous student numbers. However, through research created questionnaires and interviews,

Baquedano- López estimated that 25% of the student body is of Indigenous descent, from Otomanguean background or Mayan backgrounds. She also discussed the visibility and sociocultural integration of Yucatecan immigrants in the United States, emphasizing the role of grassroots organizations in preserving cultural practices and supporting legal and health services. The author concludes the need to continue efforts to understand the immigrant Mayan population schooling experience.

More recently, Campbell-Montalvo (2021) investigated how Indigenous Mexican languages were accounted for in a Florida school, using a mixed-method approach incorporating both qualitative and quantitative data. Campbell-Montalvo used four primary data sources: observations from school registrations, a comprehensive language survey of 1,287 students or their parents, interviews with 46 school-related individuals, and analysis of 1,956 school records. The objective was to examine the linguistic landscape, capturing data on race, ethnicity, and languages spoken at home and by parents. Findings revealed significant underrepresentation of linguistic diversity in school records, which listed only 13 languages compared to 29 identified by the authors survey. According to Campbell-Montalvo, this undercounting of Indigenous Mexican languages suggests systemic flaws rooted in racial and linguistic biases, such that language is associated with race and ethnicity, leading to misclassification of a student's home language. Campbell-Montalvo spotlights the gap between student and caregiver actual language use and what is recorded, which she claims is exacerbated by school registration processes and district policies. Despite some parents viewing the loss of their native language as a trade-off for better educational opportunities, Campbell- Montalvo stresses the need for schools to accurately reflect and support the multilingual realities of their communities, moving beyond superficial language surveys to foster true linguistic diversity and inclusivity.

In her 2023 book, Campbell- Montalvo, explains the concept of “Latinization” and its impact on the erasure of Indigenous identity in Central County, Florida schools. Latinization happens when speaking Spanish is equated with being Latinx, which consequently excludes Indigenous languages and identities. Through her multi-year ethnographic study, Campbell-Montalvo describes how Indigenous languages are systematically underrepresented, with students often misclassified as Spanish speakers based solely on their country of origin. This misclassification results in the erasure of Indigenous identity within the educational system. Additionally, Campbell-Montalvo's research uncovered significant discrepancies between school records and the actual linguistic diversity present among Indigenous students. For instance, while one school reported only one Indigenous language, Mixtec, Campbell-Montalvo's survey identified at least eight Mesoamerican Indigenous languages spoken. Similarly, while another school recorded 13 languages spoken by parents, the research revealed 29. According to Campbell- Montalvo, Indigenous languages were being omitted from official school language records by a factor of 19.

Statement of Purpose

Previous research has started to examine the linguistic diversity among Mesoamerican Indigenous children in US schools, however, this research frequently oversimplifies the description and identification of the Mesoamerican languages present in US schools. Specifically, there is broad and inaccurate grouping and homogenizing of Mesoamerican Indigenous languages into general labels such as “Mixtec,” “Mayan” or “Zapotec”. Velasco's (2010) case study illustrates this, where she provides a general overview of Indigenous languages but fails to acknowledge the distinctiveness of each language, resulting in overlooking specific details about the languages spoken by her participants. She notes Mixtec as the language used by

the two participants in her case study; however, 'Mixtec' is not a singular language. Mixtec encompasses a variety of distinct languages that can be as different to one another as English is to German or Portuguese is from Italian (DiCanio et al., 2019). Similarly, Machado-Casas (2012) identified several indigenous parents representing six different languages from Mexico (Otomi, Nahuatl, Maya, Zapoteco, Mixteco, and Tzotzil), one indigenous language from El Salvador (Pipil), and two from Guatemala (Quiché and Kaqchikel). While Quiché and Kaqchikel are labeled in a way which can be identified, using the labels, Maya, Zapoteco and Mixteco obscures and homogenizes the actual language used by caregivers (Pérez Báez, 2013; 2017). When languages are homogenized, the specific educational needs of speakers of different languages may be overlooked. For example, if a specific interpreter is needed for an initial school meeting, getting a "Mixtec" speaker without understanding that there are significant linguistic differences within the Mixtec languages spoken by the caregivers can result in inequitable access of information at that meeting. A lack of precise language information hinders both effective communication and the understanding of a families' linguistic diversity (Arauju, 2009; Baquedano-López et al., 2013; Campbell-Montalvo, 2021)

The presence of Mesoamerican culture and language diversity in the US begets the need to re-examine our current educational policies for supporting English learners, particularly in terms of accurately identifying the languages these children are exposed to facilitate the implementation culturally responsive and linguistically appropriate educational practices. Students of Mesoamerican Indigenous origin are often not recognized and are frequently lumped under broader classifications based on nationality, such as Guatemalan or Mexican, rather than being identified as a distinct community that uses languages other than Spanish at home.

Additionally, they are sometimes incorrectly categorized as non-Indigenous within broad labels like Hispanic or Latino (Batz, 2014; Campbell-Montalvo, 2023; Peñalosa, 1985; Yescas, 2010). To address these challenges, the impact of state and federal requirements on the identification and recording of these languages within educational contexts needs careful consideration. By law, children who speak languages other than English are funneled through an English Learner identification process. This process is particularly pivotal for children who use languages that are not as well-known as is the case for Mesoamerican languages in the US. It is at this stage that schools must ensure the accuracy of the linguistic information gathered, as it will significantly influence the educational support provided to these students.

This project thus seeks out to conduct a comprehensive review across all 50 states, examining student HLS data and state language codes for documenting home languages. The research focuses especially on Mesoamerican Indigenous languages. The goal is to evaluate if state practices comply with federal guidelines to identify non-English languages and to pinpoint discrepancies and gaps in these practices. This initiative builds on an exploratory analysis of seven states by Zyskind et al. (2024) which explored language codes and identification practices of established immigrant destinations like California, Texas, New York, and Florida, and newer ones like Georgia, Utah, and Oregon. In this preliminary study, we observed varied approaches to language documentation. States like Florida, Georgia, California, Oregon and Utah have developed new language codes to more inclusively represent Mesoamerican languages, with wide variation to the level of language granularity observed. Conversely, New York and Texas have adhered to the more restrictive ISO 639-2 standards, which only recognize broad language families such as Mayan, and branches like Zapotec, Otomian from the Otomanguean family, and Nahuatl from the Uto-Aztecan family. Given the diversity seen in seven states, this project aims

to expand our understanding of how SEAs record non-English languages nationwide, seeking insights that could inform more accurate language recording or guide improvements in these practices and future policy revisions.

The research questions that will be explored in this study are as follows:

1) How and to what extent do state department agencies identify and document Indigenous Mesoamerican languages among English Learners?

The second part of the study will be a case study using quantitative methods to explore the impact of the Oregon Department of Education's 2020 update to the Language of Origin code list, which included addition of 150 Indigenous Mesoamerican languages previously missing from the state language code list. The second research question is:

2) What is the immediate impact of introducing 150 detailed Indigenous Mesoamerican language codes in 2020 on the number of these languages identified and recorded by the Oregon’s SEA?

CHAPTER II
METHOD

Participants

Participants for research question one included two main groups, the five coders who engaged in structured policy review (n = 5) coding and meetings. The second group were nine representatives from State Department Agencies (SEA) who participated in a Qualtrics survey to learn more about their state language code inclusion of Mesoamerican languages (n =9)

Coders

Five coders were involved with the data collection and coding of state language practices. Coders were tasked with gathering state language documents for ten assigned states and to code the state language lists according to the established categories. Coders also attended five meetings or watched recorded meetings to engage in coding training, to meet discuss findings about their individual states, to generate new coding typology and to offer suggestions to direction of the survey that would be sent out to state department agency (SEA) representatives. The coders included three doctoral students in special education, an undergraduate student with a background in linguistics, and a graduate student in a principal licensure graduate program. Coders all completed CITI Training when IRB was required to engage with SEA representatives via Qualtrics survey.

Recruitment. Coders were recruited through and online job posting via University of Oregon Handshake and through email description of study.

Inclusion Criteria. Five coders with background in education, policy and special education were sought. Moreover, they needed to be able to commit to initial training and ongoing meetings.

Compensation. Coders were offered an hourly rate of $17 to code 10 states and to attend five meetings.

State Department Agency Representatives

Sample. A researcher-developed Qualtrics survey was distributed to representatives from 22 state education departments that included five or more Mesoamerican languages in their language code lists. Nine state department agents participated and completed the survey, representing a range of job titles such as Directors, Associate Directors, Program Specialists, and Assistant Directors. These representatives were from six states: Colorado, Washington, Oregon, Minnesota, South Carolina, and Connecticut. Demographic information about the participants was not collected, as the primary focus was to gather insights about state language codes rather than the representatives themselves. The distribution of participants by state included two representatives each from Oregon, Minnesota, and Washington, and one representative each from South Carolina, Connecticut, and Colorado.

Recruitment. State department agency representatives were recruited via email. The principal investigator (PI) identified individual email addresses or those associated with state EL programs and Title III programs. The recruitment email, created in accordance with University of Oregon guidelines and approved by the Institutional Review Board (IRB), included a brief description of the survey and a link to access it.

Inclusion Criteria. Only states that incorporated five or more Mesoamerican Indigenous languages into their language codes for documenting the home languages of English Learners (ELs) were targeted. Emails were sent to individual representatives or group addresses associated with the following states: Indiana, Washington, Minnesota, Oregon, Colorado, Illinois, Florida, North Carolina, South Carolina, and Connecticut.

Compensation. No compensation was offered. Completion of survey was voluntary.

Setting

All phases of this research project were conducted online. Public data was gathered from state department websites and entered OneDrive, a Microsoft cloud-based platform that enables users to store, access, and manage files online, as well as share them with team collaborators. Coding and ongoing meetings with coders were conducted on Zoom. Surveys were distributed to state education representatives using Qualtrics, an online survey platform designed for creating and distributing surveys, collecting responses, and analyzing data.

Measures

Public data or data acquired via public request in state department websites was the main source used to collect necessary data to answer both research questions. Moreover, a researcher created survey was used to gather necessary data to answer the first research question. The second data source to answer research question two was acquired through Oregon Public Records Request to obtain de-identified student language data. Details for each measure is provided below

Public Data

Data was systematically gathered from publicly available sources located on state education department websites. Following Baker et al. (2024), the data collection process involved a systematic review of each SEA website. The search terms used were primarily focused on locating information related to English Learners (EL), language codes, Home Language Surveys (HLS), and Every Student Succeeds Act (ESSA) state plans across various State Education Agency (SEA) websites. The overarching approach was state-centric, examining individual states to gather specific information pertinent to each state's EL language policies.

Consistent keywords such as "Home Language Survey," "ESSA State Plan," "Language Codes," and "Top 5 languages" were employed across different states to maintain uniformity and facilitate comprehensive searches. For instance, to find general information about English Learner programs and resources specific to a state like Arizona, search terms like "Arizona English Learners," "EL data," and "EL forms " were used. Similar patterns were followed for other states by adjusting the state name accordingly. When seeking ESSA state plans, phrases like "ESSA state plan Arizona" or "ESSA Plan website North Carolina" were effective in accessing each state's consolidated plan, which outlines how federal requirements, including those related to ELs, are addressed. To locate Home Language Surveys, searches such as "Home Language Survey in Arizona SEA search tab," "Home Language Survey search at CDE website," or "Home language survey Illinois Google" were conducted. These searches aimed to find the specific HLS documents used by states to identify students who might need EL services. When searching for language codes and data systems, terms like "Language Codes in CA SEA search box," "Colorado language codes in Google," and "student information system department of education" were employed. This facilitated finding the codes and classifications used by states to identify and record students' native languages in their data systems. In efforts to obtain data on the most documented non-English languages among EL students in each state, phrases like "Top 5 languages," "EL learner students by language," and "Languages Top 5 ELs" were used. This helped in collecting data on EL populations, including the prevalence of specific languages within each state. Additionally, searches for EL program resources and guidance involved terms like "EL forms tab," "English Learner Guidebook," and "English Learner Identification," which were intended to find resources, manuals, and guidelines related to EL program implementation and student identification processes. Both internal SEA website search functions and external

search engines like Google were utilized to maximize the chances of finding the required documents. For example, when information was not easily accessible through site navigation, the Google search engine was used with targeted phrases such as "Home language survey in Kansas Website" or "Language codes on Minnesota website." See samples of the data collection table in Appendix A.

Oregon Language State Language Diversity Data

After reviewing publicly available data on language counts and corresponding with state representatives, a formal request was submitted to obtain detailed records of specific languages documented in Oregon schools from 2016 to 2024. The data provided included the language name, enrollment numbers, total enrollment, languages classified within the English Learner (EL) category, total English Learners, and the percentage of EL students for each language from 2016 to 2023

Survey on Language Identification Practices for Indigenous and Multilingual Learners

State department representatives completed a survey on Qualtrics. The survey consisted of questions about state procedural information related to language codes and the documentation of Mesoamerican Indigenous languages. The questions focused on participants' roles within their State Education Agencies (SEAs), decision-making processes, and experiences with language code policy development. All questions were designed to avoid eliciting sensitive personal information, limiting responses to professional experiences and insights. The survey included questions about the participant’s role within the State Department of Education, their familiarity with language codes for documenting non-English languages spoken by English Learners, the state’s practices for including Mesoamerican Indigenous languages in language code lists, and

their perspectives on federal policies and guidelines related to native language documentation in schools. See Appendix D for example questions

Research Design

Research Question one

Building on methodologies from previous studies by Callahan et al. (2022), Umansky et al. (2022), Baker et al. (2024), and Leider et al. (2021), this project employed Critical Policy Analysis (CPA) to address research question one. The focus was on how state policies implement federal requirements for identifying non-English languages, with a particular emphasis on Indigenous Mesoamerican languages. Among CPA's fundamental concerns, this study concentrated on "Policy Rhetoric vs. Practiced Reality", examining the gap between policy intentions and their real-world implementation (Diem et al., 2014). Specifically, the study aimed to investigate how federal guidelines for non-English language identification unfold at the state level and what this means for educational equity for ELs. Moreover, structured document analysis was conducted to review official native language policy documents from state education agency websites and language identification protocols. The goal was to understand how states accounted for Mesoamerican Indigenous languages in their state language codes.

Research Question Two

To address Research Question 2—examining the immediate impact of introducing detailed Indigenous Mesoamerican language codes in 2020 on the number of these languages identified and recorded by the state—a Regression Discontinuity (RD) design was applied. RD is a quasi-experimental method commonly used to evaluate the outcomes of interventions or policy changes when random assignment is not feasible. It allows for the study of immediate outcomes related to a treatment or policy change by comparing data before and after a pre-established

cutoff point (Shadish et al., 2002; Cook & Campbell, 1979). RD has been widely used across fields such as social welfare, policy, business, and health initiatives (Shadish et al., 2002). In this study, the treatment cutoff was defined as the implementation year of the new language codes (2020). The pre-cutoff control group consisted of data from the school years 2016–2019, while the post-cutoff treatment group included data from 2021–2023. Dependent variables included (1) the frequency of specific Mesoamerican Indigenous languages identified, (2) changes in the frequency of the "Other" label, and (3) the total count of the Spanish language label before and after the policy change.

Procedures and Data Analysis

Research Question One

Phase 1. The study began by assembling a team of five coders who were trained to gather publicly available data across state department websites. The training process for coders involved several structured phases designed to ensure a standardized approach to reviewing state documents related to the documentation of non-English languages, with a specific focus on Mesoamerican Indigenous languages. Five coders were recruited and trained to systematically code and organize policy information into a standardized memo template. The training began with a 90-minute session that included a PowerPoint presentation covering foundational content on Mesoamerican languages, ISO 639-2 and ISO 639-3 standards, and federal language reporting requirements under ESSA. Coders were instructed on how to navigate State Education Agency (SEA) websites and provided with a list of search terms to locate relevant language policy data. A live demonstration was conducted to model how to search for and document data from a designated state website. Coders then spent 30 minutes independently locating and recording information on a demonstration spreadsheet to confirm their understanding.

Phase 2. The coding process involved two distinct iterations. In the first three weeks, coders conducted an initial review of 15 states, focusing on states known to have significant Mesoamerican populations based on prior research and immigration reports. This initial coding focused on documenting non-English language practices and analyzing whether states adhered to or deviated from federal guidance, specifically ISO 639-2 standards. Coders reviewed public documents, such as home language surveys, state language codes, ESSA state plans, and related policy documents, to identify practices that facilitated or hindered the documentation of Mesoamerican Indigenous languages. States were categorized based on their typology: adherence to ISO 639-2, passive differentiation (inclusion of at least one additional language), or active differentiation (inclusion of two or more additional languages or the use of ISO 639-3 codes). Coders recorded their findings in a shared Excel spreadsheet, which documented state names, URLs of sources, access dates, document types, and mentions of Mesoamerican languages.

Phase 3. The second coding iteration. Over the following four weeks, after completing the initial coding for 30 states, the coding process entered a refinement phase. Coders participated in group discussions to review and refine the typologies based on observed patterns and ensure consistency in state categorization. During this phase, inter-rater reliability was assessed using Cohen's Kappa, with a threshold of 0.60 or higher indicating acceptable agreement. Discrepancies in coding were resolved through consensus meetings, which also led to adjustments in the coding scheme. Following multiple group discussions, individual meetings with coders, and consultations with Dr. Pérez Báez, linguistic expert, it was determined that the original coding scheme required simplification to better capture nuances in state practices. The initial framework—categorizing states as demonstrating active differentiation, passive

differentiation, or no differentiation—was abandoned in favor of a binary approach. Coders re-coded specific practices using binary values (0s and 1s) to determine whether states incorporated more granular language codes, such as Kanjobal (listed in ISO 639-3), or provided language codes that went beyond both ISO 639-2 and ISO 639-3.

Phase 4. Final coding development. Forty-three state language code lists were analyzed using the new coding scheme. To gain deeper insight into the factors driving the inclusion of Mesoamerican languages and to address research question one in greater detail, a Qualtrics survey was developed to understand how states incorporated these languages into their language code lists, especially in cases where they deviated from federal guidelines to be more inclusive. The survey consisted of fourteen questions, focusing on the respondents' roles within their State Department of Education, their familiarity with language codes for documenting non-English languages spoken by English Learners, and their state's practices for including Mesoamerican Indigenous languages.

Analytical plan

The analyses were shaped by the project detailed by Umansky et al. (2022), who conducted a fifty-state policy review to examine how states across the United States identify Indigenous students as English learners. To distinguish between state practices, Umansky and the team created four typologies: no differentiation, possible differentiation, ambiguous differentiation, and clear differentiation. Following the approach of Umansky et al. (2022), this study categorized state language documentation practices into distinct patterned groups based on their approaches to documenting Mesoamerican Indigenous languages. Iterative memo-writing techniques were used to systematically categorize and analyze state policy data, emphasizing the nuances of how Indigenous languages were identified and recorded (Saldaña, 2014). Saldaña's

approach to memo-writing involves crafting narrative memos that capture the coder's interpretations, insights, and thought processes about the data. This was an iterative process in which coders had multiple opportunities to provide input on the coding and the interpretation of policy documents. Group meetings and discussions further supported the interpretation of data and the exploration of deeper meanings and patterns. State practices were categorized into typologies to represent the most common approaches, a strategy previously applied in analyzing 50-state education policy reviews (Bailey & Kelly, 2013; Umansky et al., 2022; Hazi & Rucinski, 2009).

Language documentation and inventorying, rooted in cultural anthropology and linguistics, focused on capturing essential information such as names, geographical distribution, genealogical classification, speaker count, and documentation levels. These efforts were critical for clearly identifying languages and avoiding ambiguities (Drude, 2018; Seifart et al., 2018). The field has expanded to address the need for understanding global linguistic diversity, advancing knowledge across linguistic domains such as phonetics, morphology, syntax, semantics, and supporting language preservation efforts (Hale et al., 1992; Hammarström, 2018; Seifart et al., 2018). Two primary resources for language classification—Glottolog and ISO 639-3—played a a large role in this study. The Glottolog, developed by the Max Planck Institute, is a comprehensive database cataloging both extinct and living languages (Hammarström et al., 2018; Nordhoff & Hammarström, 2011). It assigns a unique identifier, the Glottocode, to every linguistic variety, including families, languages, and dialects, and organizes them within a genealogical classification known as the Glottolog tree (Nordhoff & Hammarström, 2011; Hammarström, 2018). ISO 639-3, managed by SIL International and linked to the Ethnologue catalog, used criteria such as mutual intelligibility, shared literature, cultural identity, and

established ethnolinguistic identities to classify language varieties (Hammarström, 2015). Both Glottolog and ISO 639-3 provided largely compatible and interchangeable identifiers (Forkel & Hammarström, 2022). Accurate language documentation is essential for recognizing and preserving global linguistic diversity (Anderson et al., 2015; Blust, 2013). In the context of this study's focus on language identification practices in schools, the level of language granularity related to whether languages were accurately recorded or could be recorded based on the codes used by states. Adherence to ISO 639-2 codes often resulted in many languages being overlooked (Zyskind et al., 2024). By contrast, using ISO 639-3 codes, which categorized individual languages rather than broader language families, enhanced the understanding of the actual languages present within states, school districts, and individual schools.

Research Question Two

A regression model was constructed to compare these outcomes across the two time periods in Oregon language identification. The regression equation was as follows:

Equation: $Y_i = \beta_0 + \beta_1 \times \text{Grouped Period}_i + \beta_2 \times \text{SpanishFrequency}_i + \epsilon_i$

- Yi= Count of "other" label used for each school year
- GroupPeriodi: Binary variable where 0 represents the school years 2016-2019 and 1 represents the grouped years 2021-2023. In this case the model compares two distinct blocks of time.
- SpanishFrequencyi: numerical variable representing frequency of Spanish language reporting annually
- Ei: Error term, the unexplained variation in Yi after accounting for effects of grouped periods and the Spanish language frequency

- B0= Intercept which will be the baseline count of “other” labels used before the introduction of the new language codes.
- B1: The coefficient for the average change in the count of “other” labels dyr to the introduction of new language codes grouped from the periods of 2020-2021 school year to 2023-2024 school year.
- B2: Coefficient that shows the frequency of Spanish language documentation and use of “other” labels

The equation was implemented and analyzed using R Studio, a statistical computing and graphics software (Posit team, 2023). R studio was also used to calculate frequency of identified Mesoamerican Indigenous languages before and after the implementation of these updated language codes was analyzed. By comparing outcomes before and after the implementation of the updated language codes, this study assessed whether the policy change led to more precise documentation of the linguistic diversity of students from Mesoamerican backgrounds.

CHAPTER III

RESULTS

Research Question 1

State education agencies are required to document and report the linguistic diversity of their English learners to the U.S. Department of Education (ED) and are provided with guidelines established by the Office of Elementary and Secondary Education (OESE). These guidelines mandate the use of the ISO 639-2 Code List—a standardized language classification system managed by the U.S. Library of Congress—to ensure consistency in language reporting across educational contexts (Office of Elementary and Secondary Education, 2024)

This study explored how state practices for documenting non-English languages are influenced by these federal requirements and guidelines. The results are presented in four parts:

1) The initial inter-rater agreement (IOA) for the coding systems developed.
2) A description of the four approaches states use for language coding.
3) Relevant policy and public data findings concerning language documentation.
4) Survey results from select state department representatives.

IOA for First Coding Iteration

A random sample of 15 states (25% of sample) was selected to evaluate and refine the language coding system. Table 1 summarizes the inter-rater reliability results, measured using Cohen's Kappa (1960) across coding categories for the following states: Colorado, Missouri, Tennessee, California, Kansas, Oregon, Florida, Nebraska, Texas, Georgia, New York, Illinois, Arizona, North Carolina, and Washington. Each coder worked within an individual spreadsheet stored in their personal OneDrive folder, which syncs with Microsoft’s cloud storage service. Ratings for each category were recorded on these spreadsheets. The ratings were reviewed to

determine alignment across coders for each category and Cohen's Kappa (1960) was used to categorize the quality of the agreement.

For most categories—Full Adherence (ISO 639-2), Partial Adherence (ISO 639-2), No Adherence (ISO 639-2), and the identification of specific Mesoamerican labels—coders demonstrated perfect agreement ($\kappa = 1.00$), which was statistically significant ($p < 0.05$). The Inclusion of Languages Not in ISO 639-2 category showed substantial agreement ($\kappa = 0.75$), with some discrepancies, and this agreement was also statistically significant ($p = 0.00729$). Substantial agreement was observed for the Typology Rating category, which included the use of active, passive, or no differentiation ($\kappa = 0.87$). However, the General Mesoamerican Classification category demonstrated only slight agreement ($\kappa = 0.122$), which was not statistically significant ($p = 0.377$) (Cohen, 1960).

After obtaining these results, a meeting was convened to discuss the challenges of applying the current coding system and to review the insights revealed by the coding process. It was determined that revisions to the coding typology were necessary to better capture and understand state practices. Following multiple meetings, discussions and individual meetings with our linguistic expert, it was determined that the coding to view language documentations practices needed to be simplified. To better capture nuances in state practices, we adopted a more investigative approach to develop the final typology. The original coding scheme, which categorized states as demonstrating active differentiation, passive differentiation, or no differentiation, was abandoned in favor of re-coding specific practices using binary values (0s and 1s). This approach allowed us to distinguish between states that incorporated more granular codes, such as Kanjobal (listed in ISO 639-3) to those that provided language codes that transcended both ISO 639-2 and ISO 639-3.

Inter-Rater Reliability for Final Typology Developed

To assess the inter-rater reliability between coders on the final typology coding categories, Cohen's Kappa (1960) statistics were calculated in R studio (Posit team, 2023) using the irr package (Gamer et al., 2019), while the boot package (Canty & Ripley, 2023) was used to calculate the standard error. The categories that coders assigned a value of either 0 or 1 were as follows:

1) Adding Codes Beyond ISO 639-2
2) Uses Labels from ISO 639-3
3) Beyond 639-3
4) Uses 9999/Unknown/Unlisted

States Selected. The dataset ratings were completed by two different coders across 12 randomly selected U.S. states generated from https://www.gigacalculator.com/randomizers/random-state-generator.php . The states selected were California, Ohio, Arkansas, Louisiana, Utah, Oregon, Hawaii, New York, Mississippi, Massachusetts, Montana, Virginia, Alabama.

IOA Results. Table 2 shows the results for IOA for the final coding scheme. There was 92.31% agreement for the category "Adding Codes Beyond ISO 639-2," with a Kappa value of 0.85, indicating substantial agreement between coders. For the category "Uses Labels from ISO 639-3," there was 100% agreement, with a Kappa value of 1.00, reflecting perfect agreement. Similarly, for "Codes Beyond 639-3," there was also a Kappa value of 1.00, indicating perfect agreement. Both coders were in complete consensus regarding the use of labels from ISO 639-3 and codes beyond 639-3 across all states, with no variability (standard error of 0.00). The category "Uses 9999/Unknown/Unlisted" had an agreement rate of 84.62%, with a Kappa value

of 0.69, representing substantial agreement between coders and a standard error of 0.21 (Cohen, 1960).

Typology State Practices Results

The following section provides an overview of the four approaches states use in their code lists to identify and document Mesoamerican Indigenous languages in schools. The four state types include: (1) Only Federal Guidance States with Mesoamerican coverage, (2) Only Federal Guidance with no Mesoamerican coverage (3) Mesoamerican branches, languages or codes from ISO 639-3 (4) States Beyond ISO 639-2 and ISO 639-3. Figure 5 displays a map of the United States with different shading for each state type.

Only Federal Guidance. The largest group of states, totaling 21, relied exclusively on the federal requirement to use ISO 639-2 language codes to document the languages spoken in their schools. This guidance, provided by the U.S. Department of Education (ED), directs state agencies to use languages and codes from the ISO 639-2 Code List. The states in this group included: Alabama, Alaska, Hawaii, Idaho, Iowa, Kansas, Louisiana, Maine, Maryland, Massachusetts, Mississippi, Missouri, New Jersey, New York, North Dakota, Oklahoma, Pennsylvania, South Dakota, Texas, Wisconsin and Wyoming. Notably, these states differed in their use of the six ISO 639-2 codes relevant to documenting Mesoamerican languages. As a result, they were categorized into two subgroups: those that included at least one relevant Mesoamerican language label and those whose language lists excluded all such codes.

Only Federal Guidance States with Mesoamerican coverage. States in the only Federal Guidance typology with Mesoamerican coverage include one or more of the relevant Mesoamerican languages from the approved ISO 639-2 list (Language Codes for EdPass, 2024), including specific language families such as Mayan (myn), Otomanguean (including Zapotec

(zap) and Otomian (oto), and Nahuatl (nah) from the Uto-Aztecan family. Or include the broader language codes like Central American Indian (cai) and North American Indian (nai) that facilitate general classification of Mesoamerican languages. States in this group included: Albama, Arizona, Kansas, Maryland, Massachusetts, Mississippi, Missouri, New Jersey, New York, Oklahoma, South Dakota, Wisconsin and Wyoming.

Although these states follow federal guidelines, variations exist in their use of ISO 639-2 language lists. States such as South Dakota, New Jersey, Wisconsin, New York, Idaho, Oklahoma, Massachusetts, Missouri, North Dakota, and Maryland incorporate all six ISO 639-2 codes, including Mayan (myn), Otomanguean Zapotec (zap) and Otomian (oto), and Nahuatl (nah) from the Uto-Aztecan family, as well as broader codes like Central American Indian (cai) and North American Indian (nai). Kansas includes four of the six codes: Zapotec (zap), Mayan (myn), Central American Indian (cai), and North American Indian (nai), but omits Nahuatl (nah) and Otomian (oto). Alabama incorporates Otomian (oto), Nahuatl (nah), Mayan (myn), Central American Indian (cai), and North American Indian (nai), but excludes Zapotec (zap). Pennsylvania includes Zapotec (zap) and Mayan (myn), but omits Otomian (oto), Nahuatl (nah), Central American Indian (cai), and North American Indian (nai). Mississippi includes Central American Indian (cai) and North American Indian (nai), but does not include Zapotec (zap), Otomian (oto), Nahuatl (nah), or Mayan (myn). Maine includes Central American Indian (cai), Mayan (myn), and North American Indian (nai), but omits Zapotec (zap) and Nahuatl (nah). Arizona includes the Otomian branch with unique numerical code (86) but excludes all five other ISO 639-2 Mesoamerican codes. Wyoming includes only Zapotec (zap), omitting all other Mesoamerican language codes.

Notably, while these states use languages from ISO 639-2, there are also variations in whether they use the 3-letter ISO coding convention. Mississippi and Arizona use numerical codes that correspond to the language and its related ISO 639-2 code. On the other hand, New York, Oklahoma, Massachusetts, Missouri, Maryland, Kansas, Wyoming, Alabama, Wisconsin, South Dakota, New Jersey, and Maine use the standard 3-letter ISO coding convention.

Only Federal Guidance States without Mesoamerican coverage. States that adhere to federal guidance by using ISO 639-2 codes but exclude all relevant Mesoamerican language labels (e.g., Zapotec (zap), Otomian (oto), Nahuatl (nah), Mayan (myn), Central American Indian (cai), or North American Indian (nai)) include the following: Alaska, Hawaii, Iowa, Louisiana, and Texas Among these states, only Iowa uses a student information system that directly applies ISO 639-3 three-letter codes. In contrast, Texas, Alaska, and Louisiana rely on numerical codes mapped to the corresponding ISO 639-2 languages. Hawaii, taking a distinct approach, presents its language list by identifying each language with its English name alongside its autonym in a separate column.

States with ISO 639-3 Language Branches, Languages and Codes

The next largest group consists of states that expanded beyond the ISO 639-2 codes by using at least one language, language branch, or code from ISO 639-3. These states include Arkansas, California, Connecticut, Georgia, Illinois, Indiana, Montana, Nebraska, New Mexico, North Carolina, Ohio, Oregon, Rhode Island, South Carolina, Tennessee, Utah and Washington.

States in this group included branches within the Otomanguean family that would otherwise go unrecognized under ISO 639-2, specifically the Mixtec, Triqui, Chatino, and Oto-Pamean branches. For example, New Mexico transitioned to a new data collection system called NOVA (not an acronym) in 2024 which enables near real-time, automated data collection and

reporting from local education agencies to the state department of education (New Mexico Department of Education, 2024). The New Mexico system uses numerical codes aligned with ISO 639-2 but also includes the Mixteco branch (coded as 86), a branch not covered by ISO 639-2 but well represented in greater granularity in ISO 639-3. Similarly, Georgia incorporates the Mixtec branch in its numerical coding system (coded as 036), while South Carolina uses a distinct three-letter code (MIX) for Mixtec. North Carolina recently updated its student information system to include the Mixteco branch. Montana and California also recognize the Mixtec branch but incorrectly categorize it under the Otomian code (oto). In addition, Montana includes Pame and Jonaz in its list. Pame is part of the Oto-Pamean branch of the Otomanguean family, while Jonaz, also known as Chichimeca-Jonaz, is also part of the Oto-Pamean branch.

In addition to Mixtec, some states also added the Triqui or the Chatino branches. For example, Indiana uses its unique three-digit numerical system to include Mixteco (633), Triqui (898), and Chatino (136). North Carolina also added the (Chatino) and Amuzgo branch from the Otomangeuan family. Washington also includes both branches using a combination of two- and three-digit codes: Mixteco (539) and Triqui (580). The Oregon language list provides a more detailed approach by explicitly providing language names within the Triqui branch, including Triqui, Chicahuaxtla (5320), Triqui, Copala (5340), and Triqui, San Martín Itunyoso (5360).

Some states in this group went beyond the branch level by adding specific languages with greater granularity that are covered by ISO 639-3. For instance, South Carolina uses a unique three-letter combination to identify specific Mayan languages such as Akateko, Q’anjob’al, Poqomam, and Mam, even though broader ISO 639-2 codes like Zapotec (zap), Otomian (oto), and Nahuatl (nah) are absent. Similarly, Georgia defines "Central American Indian Languages" with specific Mayan languages, including K’iche’, Mam, and Q’eqchi’, providing greater detail

despite omitting the broader "Mayan" label (myn) from ISO 639-2. Likewise, Ohio includes ISO 639-2 codes and adds ISO 639-3 granularity for Mayan languages like K'iche', Q'eqchi', and Ixil. Indiana uses its numerical system to add several Mayan languages with greater granularity: Poqomchi (723), K'iche' (736), Mam (597), Kaqchikel (423), Ixil (412), Chuj (152), and Akateko (017). Illinois adopted a similar approach with its numerical coding system, including Achi (291), Chuj (169), Mam (191), Ixil (293), K'iche' (186), Tzotzil (277), Q'anjob'al (159), and Q'eqchi' (217). Arkansas uses two-letter codes that do not conform to ISO 639-2 or ISO 639-3 coding conventions, but it includes Mayan languages from ISO 639-3, such as Tz'utujil (TZ), Q'anjob'al (QI), and Q'eqchi' (QI). Similarly, Connecticut's language system consists of 230 three-digit numerical codes that add Mayan languages such as Chuj (237), Jakalteko (235), Q'anjob'al (228), K'iche' (214), Mam (232), and Poqomchi' (224). Lastly, North Carolina added the following Mayan languages Aguacateco, Mam, Kanjobal.

Some states included only one language from the Mayan family in their language lists. For instance, Tennessee and Utah both added K'iche' (quc) as a separate Mayan language. Similarly, Nebraska used its four-digit coding system to include Q'anjob'al (6030) as the only Mayan language. Rhode Island provides detailed granularity for K'iche' under numerical code 3731 while also using the broader ISO 639-2 label (myn) Additionally, it lists K'iche' under code 6020 with the more specific ISO 639-3 code (quc)

Oregon, in contrast, stands out with its system of 650 codes, offering unprecedented granularity by naming specific individual languages: over 50 Zapotec languages, 54 Mixtec languages, 36 Nahuatl languages, and several Mayan languages, including K'iche', Mam, and Q'anjob'al.

Lastly, the language isolate Purépecha, a language with no known linguistic relatives or connections to other Mesoamerican language family is included in the language codes of several states. Oregon provides a specific code for Purépecha, while Illinois represents it with the numerical code 212. In contrast, states like North Carolina and Washington use a broader hypernym “Tarasan” to classify Purépecha.

States Beyond ISO 639-2 and ISO 639-3:

A smaller group of states demonstrated the most unconventional approaches to language documentation, incorporating labels and languages that extend beyond the standard ISO 639-3 framework. This group included: Colorado, Florida, Minnesota, and Virginia. These states not only incorporated languages directly from ISO 639-2 or ISO 639-3 but also use distinct naming conventions and additional localized labels not present in Glottolog or in either ISO 639-2 or ISO 639-3 lists.

Colorado's language codes, detailed in the Data Pipeline Frequently Requested Codes section, combine ISO 639-2 and ISO 639-3 conventions with alternate names and labels not listed in either standard. Notably, there are language labels that refer to specific variants within the Nahuatl and Mixtec branches and within the Mayan language family. For instance, for the language Nahuatl, Eastern Huasteca (NHE) found in ISO 639-3, the code list also includes alternate names not listed in ISO 639-3, such as Eastern Huasteca Aztec, Náhuatl de Hidalgo. Similarly, languages within the Mixtec branch are listed with distinct labels such as "Eastern Juxtlahuaca Mixtec" and "Mixteco de San Juan Mixtepec," which expand the specificity of which exact Mixtec language this refers to and where it is used.

The Colorado codes also included a range of Mayan languages with distinct labels. For example, the language Kaqkchikel, Central (cak), is listed both with spelling variants like

Cakchiquel, Kaqchikel, and Kaqchiquel, Kaqkchikel, but also lists a distinct label that is not recognized in ISO- 639-3, Yepocapa Southwestern. Likewise, the Mayan language Chuj (cac) had alternate names with spelling variations Chuh, Chuhe, but also lists specific labels denoting local identifiers Chuj de San Mateo Ixtatán, Chuje, and Koti'. Similarly, the Mayan langauges Ixl (ixl), Jakalteko, Western (jai), and Q'anjob'al, Eastern (kjb), and K'iche' Central (quc) and Mam (mam) all found in ISO 639-3, had alternate names listed that pointed to specific localized labels (i.e. Awakateko-Ixil, Santa Eulalia Kanjobal, Qach'abel, Huehuetenango Mam). Many of these alternate names and labels that fell outside the ISO 639-3 list instead are likely local or community labels.

Virginia's system, documented in the "Student Record Collection Code Values," includes 740 codes that combine numerical identifiers, ISO 639-2 labels, and state-specific entries. While some languages come directly from ISO 639-2 or ISO 639-3, Virginia also employs unique naming conventions that are not found in Glottolog or either ISO. The system includes languages like Zapotec (zap) and Otomian (oto), but also features additional entries like "Mixteco, San Juan Mixtepec" and ISO 639-3 Deprecated Mayan varieties, such as "Central and Eastern Cakchiquel," and region-specific Quiché labels like "Cunen Quiché" are also included, reflecting Virginia's integration of local and unique naming conventions.

Minnesota's MARSS system lists 442 language codes, ranging from one to three digits. The system includes standard ISO 639-2 codes like Zapotec (402) and Nahuatl (154) but also uses alternate names like "Eastern Huasteca Aztec" and "nāhuatlahtōlli," which fall outside ISO conventions. Specific Mayan languages, including Kaqchikel (439) and Jakalteko (442), are listed with localized names such as "abxubal" and "Jab' xub'al," which are absent from both ISO 639-3 and Glottolog. Similarly, Mam (425) and Tojolabal (441) are labeled with alternate names

such as "Qyol Mam" and "Comiteco," further highlighting the use of distinct identifiers not found in other language standards.

While other states in this group added multiple languages or language identifiers not found in ISO 639-3, Florida also added one unique identifier that placed it in this typology. Florida's language codes, documented in "Appendix N: Language Codes," use two-letter combinations for 386 languages, including 17 Mesoamerican labels. Notably, Florida includes Chalchiteco (CR), which is absent from ISO 639-3 and Glottolog,

States Missing and Excluded from Typology

Seven states did not provide their language code lists. Nevada, Michigan, and Kentucky did not have their lists readily available, and state representatives did not respond to inquiries. Delaware, Vermont, West Virginia, and New Hampshire responded but did not provide lists, citing reasons such as not having specific lists, using ISO 639-2, or stating that their lists are uniform to meet U.S. Department of Education requirements for the Consolidated State Performance Reports (CSPR).

Table 3 Provides the descriptives and frequency of the Mesoamerican languages found across the 43 states language codes. Table 4 shows the number of Mesoamerican languages found in the 43 states reviewed. Figure 6 shows the most frequently identified languages across all 43 states.

Policy Considerations for States Featuring Mesoamerican Languages

In addition to examining language codes, specific language policies were reviewed to understand how languages are accounted for in state guidelines and whether Mesoamerican languages were present in publicly available student data.

Under ESSA (2015), states are required to define what it means for a language to be present "to a significant extent," ensuring the inclusion of the most common non-English languages in their definitions and aligning with statutory requirements to provide native language assessments. However, states vary widely in their interpretations of this requirement. An overview of state practices reveals that South Carolina sets a threshold at 1.5% of the total English Learner (EL) population, recognizing languages such as Spanish, Russian, Vietnamese, Chinese, and Arabic. Alabama defines significant languages as those spoken by 2% or more of the student population. Georgia sets the threshold of significant languages as those spoken by 3.0% or more of the student population. In contrast, Ohio requires a language to constitute at least 20% of the EL population, with Spanish being the only language meeting this criterion. New Mexico identifies significant languages as those representing 10% of the tested population. States such as Rhode Island, New Hampshire, Missouri, Kansas, Maryland, Utah, and Virginia apply the Office of Civil Rights' recommended threshold of 5%, though Spanish is currently the only language meeting this criterion. Tennessee defines significant languages as the five most commonly spoken home languages of ELs that meet one of three criteria: (1) 4% of the overall student population, (2) 20% of the population within a single LEA, or (3) 20% within a single grade level statewide. Colorado sets its threshold at 5% or 1,000 individuals (whichever is less) within the state's grade-level EL population, while Florida considers languages spoken by more than 5% of the student population significant for reporting. Minnesota, Arkansas, and California set their definition of significant non-English languages by pinpointing the most prevalent languages used by students in their states. In Minnesota this means determining the three most populous languages other than English identified via the home language survey. In Arkansas and California this involves identifying a language spoken by more than 15% of the student

population. Notably, despite the variation and differing definitions of “significant” non-English languages among states, the source for identifying these languages is the language codes lists that are acceptable to the ED via EDPass (ED, 2024).

The diversity of languages identified in state policy documents or publicly available data also varies widely across the United States. Spanish remains the dominant language among ELs in most states. In contrast, Ohio and Indiana report their top EL populations are Somali, Arabic, and Burmese. Montana and New Mexico highlight Indigenous languages, including Hutterite German, Navajo, and Keres, while Minnesota underscores the linguistic diversity among refugee communities, including languages such as Cushitic, Hmong, and Karen. Regarding Mesoamerican languages, Illinois reports that the top five languages among ELs are Spanish, Arabic, Polish, Urdu, and Russian, with Kanjobal—a Mesoamerican language—ranking among the top 50 native languages spoken by ELs during the 2019-2020 school year. Alabama, while not including specific Mesoamerican language labels in its official code list, ranked "Central American Indian" as the second most used language among ELs and "Mayan languages" as the fifth during the 2019-2020 school year. Similarly, in Nebraska, where only one Mesoamerican language label (Q’anjoba’l) is used, "Mayan languages" ranked as the fifth most common language spoken by ELs during the 2021-2022 school year. Rhode Island, which includes only one Mesoamerican language label (Quiche), maintains a public language dashboard for ELs. For the 2022-2023 school year, the recorded counts were 66 for Mayan languages and 143 for Quiche, with the top five languages being Spanish, Creole and Pidgins (Portuguese-based), Portuguese, Arabic, and Quiche. California also provides publicly available data and during the 2023-2024 school year, the top five languages in California were Spanish, Mandarin, Vietnamese, Arabic, and Russian. Regarding data for Mesoamerican languages, the state

recorded 191 Otomian (Mixteco), 32 for Mayan languages, and 8 for Zapotec. In Florida, "Mayan languages" ranked among the top five EL languages in the 2019-2020 school year, with 3,050 speakers recorded, alongside Spanish, Haitian Creole, Portuguese, and Arabic.

Survey results

To better understand the factors driving the addition of Mesoamerican languages and to address research question one in greater detail, a Qualtrics survey was sent to representatives from twenty-two state education departments that included five or more Mesoamerican languages in their language code lists. Contact information for individual state department representatives associated with English Learner and Title III departments was used whenever available. This included two representatives from Ohio, South Carolina, Washington, Georgia, and Florida, as well as a representative each from Rhode Island and Connecticut. When specific contacts could not be identified, the survey was sent broadly to group emails associated with English learner or Title III departments. Group emails were sent for Illinois, South Carolina, Virginia, and Minnesota.

The survey consisted of fourteen questions, focusing on the respondents' roles within their State Department of Education, their familiarity with language codes for documenting non-English languages spoken by English Learners, and their state's practices for including Mesoamerican Indigenous languages in their language code lists.

Nine surveys were completed fully, one partially completed, and one was started but not submitted. Of the nine fully completed surveys, two were from Oregon, two from Washington, one from Colorado, two from Minnesota, one from South Carolina, and one from Connecticut. Respondents represented various positions, including Directors, Associate Directors, Program Specialists, and Assistant Directors from the following six states: Colorado, Washington,

Oregon, Minnesota, South Carolina and Connecticut. See Table 5 for an overview of state representatives that responded and associated job titles. The respondents provided insights on their familiarity with language lists, policies for adding new languages, and partnerships with relevant communities.

Roles and Familiarity with State Language Codes

Respondents were asked to indicate their level of familiarity with their state's language codes, ranging from "Somewhat familiar" to "Very familiar." Program specialists from South Carolina and Minnesota, as well as directors from Colorado, Oregon, and Connecticut, reported being "Very familiar" with their state's language codes. Additionally, a program specialist from Oregon indicated being "Somewhat familiar." In Washington, an Assistant Director and an individual with an unspecified "other" role reported lower familiarity, describing themselves as "Slightly familiar" or "Somewhat familiar" with their state's language codes.

Policies for Adding Languages

There were notable differences in what state representatives identified as leading to the inclusion of languages outside of ISO 639-2 labels. In Minnesota, a program specialist wrote a comment about the impact of transitioning from a two-digit to a three-digit language code system, which increased the number of identifiable languages and enhanced the inclusivity of the state's language list. According to this specialist, this transition has enabled Minnesota to recognize over 440 languages. In Washington, the associate director explained that districts can request the addition of new languages at any time by contacting the Student Information team, with these requests prioritized for the next update. However, the process previously lacked input from language experts, resulting in issues such as duplicate entries (e.g., Q'anjob'al and Kanjobal). To address these concerns, Washington is developing a more consistent and expert-

informed review system, but specific information about when this update is set to be enacted was not provided. Meanwhile, Colorado, Oregon, and Connecticut reported having procedures in place for adding or modifying language codes but did not provide specific details about their processes.

Actors Involved in Decisions

The decision to add new languages often involved multiple actors, including state education agency leadership, local school districts, and, in some cases, parents and students. Oregon state reps highlighted collaborations with researchers from NILI (Northwest Indian Language Institue) the University of Oregon, and Oregon Tribes as instrumental in this process. In Washington, local school districts and the state agency played key roles, with additional input from language experts to develop a more uniform system and avoid duplicate language codes. The Colorado state representative reported that the addition of Mesoamerican languages resulted from efforts led by state education agency leadership, local school districts, students, and parents. In South Carolina, alongside local school districts, the South Carolina_Department of Education Title III/ML office and the Office of Research and Data Analysis were also involved. Minnesota indicated that changes to their language codes were driven by state education agency leadership, local school districts, and federal agencies. Finally, in Connecticut, in addition to state education agency leadership, non-leadership SEA staff were also involved in the process.

Student Information Systems and Editing Capabilities

The student information systems used in Washington (CEDARS), South Carolina (PowerSchool), and Colorado (Data Pipeline) allow for edits to their language lists. Similarly, Minnesota's system, MARSS (Minnesota Automated Reporting Student System), also supports modifications. In Oregon, a program specialist reported that multiple student information

systems are in use, including Synergy and OMSIS, but that there are edits allowed to be made. In contrast, Connecticut indicated that their system, the Public-School Information System (PSIS), does not permit editing to add languages.

Plans to expand or refine the language codes

When asked if their state had plans to refine or expand their code lists, the following states responded affirmatively Colorado, Washington, Oregon, South Carolina and Connecticut.

Familiarity with Federal Guidelines

When state education representatives were asked if federal guidelines and policies for adding languages to their codes were easy and clear to follow, representatives varied in their responses. Several inquired if the guidelines referred to the ISO standards, while others indicated unfamiliarity with the federal guidance. In Oregon, one of the state representatives noted that while they can provide feedback on proposed changes, they have no authority over federal codes. Instead, state codes are approved by the ODE Data Governance Committee without requiring federal approval. The South Carolina representative indicated that federal guidelines are unclear. The Minnesota representative highlighted challenges arising from the lack of adoption of ISO 639-3 language codes for federal reporting. According to Minnesota, while the state captures all home languages based on parent input, these often must be mapped back to broader ISO 639-2 language codes, which lack the specificity needed to represent certain languages. For example, a reported home language of "Kutin" is not included in ISO 639-2, requiring it to be categorized under a broader language family.

Capturing Linguistic Diversity in their State

State representatives shared varying perspectives on how their language lists and resources capture linguistic diversity. In Colorado, a representative reported confidence in their

system, stating, "We document all languages that are identified and spoken by students in Colorado." Similarly, Minnesota emphasized their proactive support for districts, explaining, "We provide resources for districts, and if a district is unsure about a language, they can contact the Minnesota Department of Education for assistance. We also identify the regions where a language is spoken and provide alternate spellings or names to simplify the process." In Washington, opinions were mixed: one representative strongly agreed that their language codes accurately reflect non-English language diversity, while another disagreed. Oregon representatives also disagreed, whereas South Carolina agreed that their system adequately captures linguistic diversity.

Documenting and support for documenting Mesoamerican Languages

When it comes to documenting Mesoamerican Indigenous languages, states expressed varying levels of confidence. Colorado reported success in capturing these languages, and Washington representatives agreed that sufficient resources are in place to support this effort. In contrast, Oregon acknowledged gaps, stating that their guidance is insufficient for accurately documenting Mesoamerican languages. Minnesota reiterated their proactive approach, providing district-level support and resources to identify languages by region and alternate names. Connecticut also expressed confidence in having adequate resources to document Mesoamerican languages.

Research Question 2

Descriptive Data

A descriptive summary of the data was completed to provide context for the analysis. From 2015 to 2023, the average number of languages reported was 273 ($M = 272.88$, $SD = 45.44$), ranging from 216 to 338. The mean percentage of students enrolled with identified non-

English languages was 36.6 % (SD = 4.86). The number of languages recorded for enrollment and for EL languages is highly skewed. Enrollment numbers vary greatly, with some languages having low enrollment counts (as low as one student) and others with enrollments of upward of 100,000. The presence of extreme values of enrollments leads to means and standard deviations that are pulled in the direction of the high outliers and as such are not meaningful to interpret. As a result, reporting the median and the interquartile range (IQR) will be used to view and understand the variability of the language data without being heavily influenced by the extreme outliers. The median enrollment for students per language was six (IQR = 59.25), with enrollment varying greatly, ranging from languages with only one student enrolled to as high as 458, 438 students enrolled for a different language. The median number of students identified as English Learners from within those identified languages per language annually was 7 (IQR = 43.00), with values ranging from 1 to 47,189 across different languages. Table 6 provides the median and interquartile range (IQR) for student language enrollment and English Learners across all recorded languages in Oregon from 2015-2023. The values represent the median count of students enrolled with a specific language as well as the median count of ELs identified for each language, calculated annually.

State Language Diversity

Enrollment Trends for Top 10 Languages from 2015-2023. For ease of reporting, when a year is mentioned (e.g., 2016) it refers to the 2015-2016 academic year and this end-year reporting will be used for all academic years from 2015-2023 (e.g., 2023 refers to 2022-2023 academic year). Among the top 10 languages reported in enrollment for Oregon students, English consistently had the highest enrollment over the years, ranging from 441,027 in 2016 to 431,842 in 2023. Spanish was the second most prevalent language, ranging from 84,627 in 2016

to 82,347 in 2023. The third most reported language from 2015-2023 was Russian, with enrollments ranging from 4,795 in 2016 to 3,288 in 2023. Vietnamese followed in fourth across the years, with enrollments ranging from 4,457 in 2016 to 3,483 in 2023. Chinese ranked fifth in spring enrollments which ranged from 3,176 in 2016 to 3,505 in 2023. Arabic, Korean, and Somali ranked sixth, seventh, and eight, respectively, each demonstrated spring enrollments that fluctuated year to year. Arabic enrollment ranged from 1,633 in 2016 to 1,703 in 2023, while Somali enrollment ranged from 1,183 in 2016 to 955 in 2023. Languages such as Chuukese and Ukrainian appeared in later years, with Chuukese enrollment reaching 1,083 in 2023 and Ukrainian enrollment reaching 1,113 in 2023. See Figure 4.

For EL the Top 10 Languages for 2015-2023. From 2015 to 2023, the top 10 languages consistently identified among English Learners (ELs) in Oregon schools were Arabic, Chinese, Chuukese, English, Other Languages, Russian, Somali, Spanish, Ukrainian, and Vietnamese. Spanish accounted for the highest number of languages within ELs, with 43,553 students in the 2016 school year with the highest enrollment recorded in 2023 at 47,189 students. Russian EL enrollment consistently ranked second across 2015-2023 school years, with the highest enrollment at 1,959 students in the 2017. Vietnamese EL enrollment reached its highest level in 2022 with 1,292 students, followed by a slight decline to 1,199 students in 2023. Chinese EL enrollment increased from 1,134 students in 2020 to 1,231 in 2023. Arabic EL enrollment was highest at 1,237 students in 2017, then decreased to 890 in 2023. The "Other Languages" category, encompassing unspecified languages, increased in enrollment from 564 students in 2016 to 829 in 2023. Chuukese EL enrollment was 865 students in 2023, up from 556 in 2016. Ukrainian EL enrollment was 797 students recorded in 2023. Somali EL enrollment peaked at 893 students in 2017 but declined to 616 students in 2021. A subset of students with English as

their primary language identified as ELs saw a gradual decrease in enrollment, from 853 students in 2016 to 598 in 2023.

Prior to Policy Change: Mesoamerican Languages Labels Use Enrollment. As seen in Table #7 The enrollment counts for total enrollment for use of the ISO 639-2 broad language labels from 2015 to 2023 were calculated. The Central American Indian (Other) category showed consistent low enrollment, ranging from one to two students annually. Two students were enrolled with Central American Indian in 2018, 2020, 2021 and only one student was enrolled with the Central American Indian label in three academic years 2019, 2022, 2023. No students were recorded with the category Nahuatl until one student was recorded in 2018 and then five, four, and six in the subsequent years until 2023. For the North American Indian (Other) category, ninety-three students were recorded in 2016 with slight decrease to eighty-six students in 2018, raised to ninety-eight students in 2019 and declined back down to eighty-nine in 2023. The Zapotec category demonstrated a decline in enrollment from 11 students in 2016 to 4 students in 2023. The Mayan Languages label was 128 in 2016, 177 in 2017, 297 in 2018, and 300 in 2019. The count for Mayan languages label began to decrease in 2021 to 296 followed by 258 in 2022 and most recently to 113 in 2023.

Prior to policy change EL Mesoamerican Broad Label Use. As seen in Table #7 The Mayan Languages language category within ELs increased from 128 in 2016 to a peak of 306 in 2019 and then decreased to 110 in 2023. The Central American Indian (Other) category showed restricted use, with counts ranging from one in 2023 to two across multiple years: 2018, 2019, 2020, 2021, 2022. The language code for North American Indian (Other) showed a decrease over time from 2015-2023, with counts ranging from forty-five in 2016 to thirty-five in 2022 and most recently thirty-eight in 2023. The Zapotec and Nahuatl categories both had consistent low

counts across the years. Zapotec ranged from zero in 2018 to six 2016 with the most recent count being three in 2023. Nahuatl counts fluctuated between one in 2018, 2020, and 2021 and two in 2019, and 2022 and most recently three in 2023.

Post Policy Change Enrollment Mesoamerican Language Labels. What follows is a summary of Tables 8-11 the enrollment frequency counts of specific Mesoamerican language labels following the language code update policy.

Zapotec Specific Language Label Enrollment. Cajonos Zapotec was identified and reported across three school years: 2021, 2022, and 2023, with an enrollment of one student per year. In 2023, Chichicapan Zapotec, Guevea De Humboldt Zapotec, San Vicente Coatlán Zapotec, and Santiago Xanica Zapotec were each identified with an enrollment of one student. Lapaguía-Guivini Zapotec was recorded in 2021 with an enrollment of one student. Santa Catarina Albarradas Zapotec and Southern Rincon Zapotec were each reported in 2022 with one student enrolled, while Southern Rincon Zapote was also reported in 2023 with one student.

Nahuatl Specific Language Label Enrollment. In 2023, Classical Nahuatl was reported with an enrollment of one student. Isthmus-Cosoleacaque Nahuatl, was recorded with one student from 2021- 2023. Isthmus-Pajapan Nahuatl had an enrollment of one student in 2022, but no students were recorded in 2023. In 2021 the Nahuatl, Orizaba language was identified with one enrollment.

Mixtec Specific Language Label Enrollment. Alcozauca Mixtec was reported in both 2022 and 2023, with one student enrolled each year. Cacaloxtepec Mixtec was reported in 2023, with one student enrolled. Diuxi-Tilantongo Mixtec were consistently reported in 2022 and 2023, with two students enrolled each year. Ixtayutla Mixtec was reported in 2021, with one student enrolled. Northwest Oaxaca Mixtec was reported in 2023, with one student enrolled. San Juan

Teita Mixtec was reported in both 2022 and 2023, with one student enrolled each year. Santa Lucía Monteverde Mixtec showed an increase from one student in 2022 to three students in 2023. Southern Puebla Mixtec was reported in 2022, with one student enrolled. San Juan Mixtepec Mixteco showed variation in identification across years, with three students in 2020, seven students in 2021, six students in 2022, and seven students in 2023.

Mayan Specific Language Label Enrollment. Aguacateco was identified in 2021 with an enrollment of two students, four students in 2022 and three students in 2023. Cakchiquel Mam was recorded in both the 2022 and 2023 academic years, with an enrollment of one student each year. Chuj enrollment began with one student in 2020, rising to six students in 2021, fifteen students in 2022, and reaching twenty-four students in 2023. K'iche' enrollment started with one student in 2020, increasing to three students in 2021, six students in 2022, and sixteen students in 2023. Mam was consistently identified with growth over the years, starting with 125 students in 2020, increasing to 161 students in 2021, 205 students in 2022, and reaching 349 students in 2023. Popti' was initially identified in 2020 with no enrollment, increasing to one student in 2021, and stabilizing at three students in both 2022 and 2023. Finally, Q'anjob'al started with two students in both 2020 and 2021, increased to 16 students in 2022, and further rose to 44 students in 2023.

Isolate Purepecha Specific Language Label Enrollment. In 2020, no students enrolled were identified as speaking Purepecha. In 2021 three students were identified enrolled under Purepecha. In 2022 and 2023, the number of identified Purepecha students remained consistent at three students per year.

***Triqui* Specific Language Label Enrollment**. Enrollment for Triqui languages increased steadily from zero in 2020 to six students in 2023. In 2021 two Triqui languages were

documented for the first time: Copala Triqui and San Martín Itunyoso Triqui, each with one student enrolled. In 2022 Copala Triqui remained at one student. Identification of San Martín Itunyoso Triqui, increased to two students. In 2023, a new Triqui language, Chicahuaxtla Triqui, was identified, with one student enrolled.

Post Policy Change EL Specific Mesoamerican Language Labels Use. What follows is a summary of Table ## showing the frequency counts of specific Mesoamerican language labels identified within the EL population following the language code update policy.

EL Specific Zapotec Language Label Use. Between the timeframe 2020-2023 only Chichicapan Zapotec was identified as a language among English Learners.

EL Specific Nahuatl Language Label Use. Only Isthmus-Pajapan Nahuatl was identified within the EL category in 2022 and 2023.

EL Specific Mixtec Language Label Use. San Juan Mixtepec Mixteco was identified within seven ELs from 2020 and 2021, four in 2022, and six in 2023. In 2023, Alcozauca Mixtec and Santa Lucía Monteverde Mixtec were each identified with one student.

EL Specific Mayan Language Label Use. Five granular Mayan languages were identified from 2019-2023 including Aguacateco, Chuj, K'iche', Mam, and Q'anjob'al. Mam was the Mayan language most identified within ELs at 165 in 2020, 214 in 2022, and 369 in 2023. Chuj was identified within the ELs category at nine in 2020, nine in 2021, 17 in 2022 and 31 in 2023. K'iche was identified within the EL category at two in 2020 and 2021, nine in 2022 and 18 in 2023. Q'anjob'al was identified within the EL category at two in 2020 and 2021, 18 in 2022 and 49 in 2023. Lastly, Aguacateco was first identified in 2022 with one EL student and two EL students in 2023.

EL Isolate Purepecha Specific Label Use. The isolate Purepecha was recorded within the EL population for one student in 2020, 2021, 2022 and 2023.

El Triqui Specific Label Use. Copala Triqui was identified in 2020 and 2021 with one student, no students in 2022 and four students in 2022-2023. San Martín Itunyoso Triqui, was identified with one student in 2022, but no students in other years.

Correlation

A correlation analysis was conducted to explore the relationship and associations between variables such as “Other Language” Grouped period (pre- and post- policy periods), Spanish enrollment prior to running the Regression Discontinuity model. The correlation matrix is presented in Table 12. The analysis included Spanish Enrollment, Spanish ELL, Other Language Count, and Grouped Period. Spanish Enrollment and Spanish ELL were negatively correlated (r = -0.77, p= *0.027*), indicating that an increase in Spanish ELL enrollment is associated with a decrease in the overall Spanish enrollment count and this was statistically significant. Spanish Enrollment was positively correlated with Other Language Count (r = 0.69, p = *0.88*), suggesting that as Spanish enrollment increases, there is also an increase in the count of students categorized under "Other Language” but this was not statistically significant. The correlation between Grouped Period (pre- and post-policy periods) and Spanish ELL was r = 0.31, p = *0.45*, which was positive but not statistically significant.

Regression Discontinuity

The primary objective of Research Question 2 was to evaluate whether the introduction of more detailed language codes—including granular Mesoamerican languages during the 2019–2020 school year—led to improvements in language documentation and changes in enrollment patterns. To understand how the identification of languages has evolved over time, a Regression

Discontinuity (RD) analysis was conducted focusing on specific language data: Spanish language Enrollment count and "Other" Language Label Count.

By focusing on a narrow window around the policy implementation, the model aims to determine if there were significant changes in language identification attributable to the policy change rather than other factors. Thus, a counterfactual comparison of language documentation practices with and without the availability of detailed language codes using the RD design was conducted with results below. See Table 13 for model results and Figure 7 for graph.

The regression equation is specified as:

Equation: $Y_i = \beta_0 + \beta_1 \times \text{Grouped Period}_i + \beta_2 \times \text{SpanishFrequency}_i + \epsilon_i$

Using the proportion of Spanish enrollment in the Spring as a predictor variable, a regression discontinuity (RD) was run to determine if the addition of granular Mesoamerican language codes led to a reduction in the use of the broad “Other languages” designation. The RD model's results indicate that there was no statistically significant change in the proportion of the "Other" language label after the policy was put into place *($p > 0.05$)*. The average proportion of the “other” language label decreased by 0.00015 or 0.015 percent during the post-policy period, but this change was not statistically significant ($p = 0.25$). Additionally, for every unit increase in the proportion of Spanish-enrolled students, an increase of 0.071 was observed, but this was also not statistically significant ($p = 0.22$).

A post hoc power analysis was conducted to assess the statistical power of the regression discontinuity model for state-level language enrollment data from 2016-2023. To estimate the observed effect size, the “pwr” package was used in R studio (Champely et al., 2020). The effect size ($f^2 = 0.130$) was derived from $R^2 = 0.1153$, with a significance level set at $\alpha = 0.05$. The results indicated a power of 0.193 (19.3%), which is below 0.80. This low power suggests that

the model had a limited capacity to detect small or moderate effects of the policy change on language enrollment trends.

CHAPTER IV

DISCUSSION

An increasing number of children in the United States are learning more than one language as they acquire core academic content, shaped by their experiences at home and school (National Center for Education Statistics, 2024). For some of these children, strong proficiency in both English and their home language is evident while others have varying abilities across their two or more languages (Castilla-Earls et al., 2019). However, the identification of another language used at home initiates a process to determine whether they are subject to federal regulations and civil rights protections that ensure they have access to grade-level curricula and gain English proficiency (Castañeda v Pickard, 1981; Lau v. Nichols, 1974). Among these multilingual learners, some will be classified as English Learners (ELs), with Spanish being the most commonly recorded home language (Office of English Language Acquisition [OELA], 2022; Park et al., 2018; Zong & Batalova, 2015).

The primary policy guiding EL education is Every Student Succeeds Act (ESSA, 2015). ESSA defines ELs as students aged 3 to 21 who use a language other than English at home and have not yet achieved proficiency in English. This group includes students who were not born in the United States, whose native language is not English, or whose environment significantly impacts their English proficiency. Under ESSA, states are mandated to establish standardized criteria for identifying ELs, implement educational programs to support their language development, and monitor their progress through assessments, educational standards, and teacher quality. Additionally, education agencies must identify the most prevalent non-English language(s) spoken by ELs to determine the need for native language assessments. Despite these mandates, state discretion in identifying and reclassifying ELs has led to immense variability in

how these policies are implemented nationwide (Callahan et al., 2022; Umansky & Porter, 2020; Villegas & Poma, 2020). As the proportion of students identified as ELs rises, it is increasingly important to understand how federal policies are translated into practice at the state level.

This study is one of the first to examine how states implement federal guidelines for identifying non-English languages beyond the home language survey, specifically focusing on discrepancies and gaps in statewide language codes and processes used to document students' language backgrounds, with a focus on Mesoamerican languages. Although linguistic diversity exists in other parts of the world, the Mesoamerican case is especially instructive due to the concentrated presence of its languages in specific U.S. regions and the increasing migration from these areas (Obinna, 2023). Recent data from Pérez-Báez (2024) signals substantial presence of Mesoamerican Indigenous languages, such as languages from the Mixtec branch, Zapotec branch, and Mayan languages such as K'iche', in areas like Woodburn, Oregon, and neighborhoods in New York City and California. Migration studies have also identified key destinations for Mesoamerican Indigenous groups, including Triqui migrants in northeastern Washington and central California, as well as P'urépecha, Nahua, Mixtec, and Zapotec communities in North Carolina, New York, and California (Holmes, 2012; Díaz de León-Martínez et al., 2020; Asad & Hwang, 2019; Kearney, 2000). Despite this documented presence, it remains unclear whether these languages and communities are sufficiently represented in schools or if current state-level language identification infrastructure is set up to accurately capture this diversity.

This study builds on an exploratory study by Zyskind et al. (2024), which found varied approaches to language documentation across eight state language codes, suggesting that some states solely rely on federal guidance which may obscure Mesoamerican languages in schools.

The goal of this project was to expand that work by conducting an overview of 50 states and to assess how states account for the growing linguistic diversity among students in U.S. schools and to emphasize the critical need for accurately understanding the specific linguistic diversity within their populations. By focusing on Mesoamerican languages, this study aims to provide a unique perspective on the gaps and variations in language identification systems set in place, especially for languages that risk being obscured or misrepresented due to current language identification procedures.

The first phase of the study used an iterative and structured approach to gather publicly available policy documents from state education department websites, federal policy documents, and language codes—to identify states' approaches to identify and document their state's linguistic diversity, with a particular focus on Mesoamerican languages. After determining that the frequency of Mesoamerican language codes indicated some states were making additions to Mesoamerican languages, survey data was gathered to further probe how these languages were incorporated. The second part of this study used quantitative methods to explore the immediate impact of introducing detailed Indigenous Mesoamerican language codes in 2020 in Oregon as a case study, assessing whether state policy changes influenced broader language documentation practices in schools. The results indicate that while there were marginal changes, these did not reach statistical significance, raising questions about supporting schools in implementing new language polices. Overall, this study uncovers flaws in current federal guidance and the diverse state practices that have emerged because of this ambiguity and the need for implementation support to ensure languages are adequately identified in schools.

The following sections present the results related to the two research questions, discuss the study's limitations, and consider implications for future research and policy practice.

Umansky and Porter's conceptual framework (2020) emphasizes the state's role in defining EL services, including teacher preparation, identification, and reclassification policies. Although this framework emphasizes the importance of recognizing language diversity as an asset in our schools, it has some assumptions built into it that this study identifies, implication will be presented.

RQ1: How and to what extent do state department agencies identify and document Indigenous Mesoamerican languages among English Learners?

Under ESEA section 1111(b)(2)(F), states must identify languages significantly represented among students and follow specific reporting guidelines. This includes using the Department of Education's (ED) list of languages for documenting non-English home or primary languages for EL students (2024). Data is submitted via EDPass, an electronic system that facilitates efficient transmission from SEAs to the Department of Education using ISO 639-2 language codes (Library of Congress, 2024). This study examined EL policy documents—including ESSA plans, state language codes, and home language surveys—to determine how each state is set up to identify and document Mesoamerican Indigenous languages. The findings indicate that state interpretations of federal guidance fall into four typologies related to their practices: (1) Federal Guidance States with Mesoamerican coverage; (2) Federal Guidance with no Mesoamerican coverage; (3) Mesoamerican branches, languages or codes from ISO 639-3; and (4) States beyond ISO 639-2 and ISO 639-3.

States' approaches to incorporating Mesoamerican language codes fall into the abovementioned typologies, spanning both established immigration states and newer destination states (Gleeson & Bada, 2019; Dondero & Muller, 2012). First, some states follow federal guidance but do not include any relevant Mesoamerican codes from ISO 639-2; this group

includes high-immigration states like Texas and newer destinations like Iowa (Sugarman, 2016). Second, states that adhere to ISO 639-2 and include some Mesoamerican languages encompass traditional high-immigrant destinations like New York and new immigration states such as Mississippi and Kansas. Third, states that expand beyond ISO 639-2 by incorporating codes from ISO 639-3—providing greater specificity in language listings—feature high-immigrant and EL population states like California and newer destinations like Oregon. These states recognize branches not present in ISO 639-2 within the Otomanguean family, such as Mixtec, Triqui, and Chatino, or include specific Mesoamerican languages from the Mayan family. The final typology consists of states that go beyond both ISO 639-2 and ISO 639-3, using distinct naming conventions and localized labels not found in ISO standards; this group includes high-immigration and EL population states like Florida and newer destinations such as Minnesota and Virginia.

There are notable issues with states that do not include any relevant Mesoamerican languages or codes from ISO 639-2, especially when there is evidence of Mesoamerican immigration to these states. For instance, Texas and Iowa currently lack codes to account for the presence of Mesoamerican languages in schools, despite growing Mesoamerican population presence. A recent study by Rodriguez et al. (2023) indicated that among immigrant dairy workers in the Texas Panhandle and South Plains, about 7.8% identified as Guatemalan nationals, many from areas where Indigenous languages are used, specifically Quiché, Huehuetenango, and Totonicapán. Iowa state codes only include the broad ISO 639-2 label (myn), despite past research demonstrates a notable presence of Guatemalan immigrants in Iowa. For instance, there is growing Guatemalan Mayan community in Dubuque, with most arrivals since 2010 speaking languages like Ixil, K’iche’, and Q’anjab’al. Many in this population arrived

as unaccompanied minors (Community Foundation of Greater Dubuque, 2024). Kaqchikel was also found to be used by meatpacking workers in Postville, Iowa, with Sandoval indicating that their children were enrolled in the local schools (Reynolds, 2013; Sandoval, 2014). As such, the state is not accurately set up to account for the linguistic diversity of students attending its schools, leading to misidentification of their language needs. Consequently, the quality of the language support for these misidentified students is questionable at best (Brooks, 2018; Menken et al., 2012)

To illustrate the limitations of relying solely on ISO 639-2 codes or restricting expansion to ISO 639-3 language branches for identifying Mesoamerican languages, California and Kansas serve as examples. California, with the highest number of English Learners (ELs), and Kansas, a newer destination state for Mesoamerican immigrants, illustrates the gaps in language documentation practices. California has the largest number of ELs in the country, with more than 20% of students in kindergarten through grade twelve designated as ELs and over 60 language groups represented (CDE Dataquest, 2014b). Despite this diversity, there is a clear lack of infrastructure to accurately capture the full range of languages spoken. California uses ISO 639-2 codes for broad categorizations and includes some granularity at the Mixtec branch level using ISO 639-3. However, the Mixtec branch is incorrectly labeled as part of the Otomian branch, when it is in fact its own distinct branch consisting of over 53 languages (Josserand, 1983). This misclassification shows some of the challenges of states attempting to extend beyond ISO 639-2 without guidance from language experts. Nevertheless, including the Mixtec branch does at least acknowledge a group that would otherwise be completely omitted if only relying on ISO 639-2 codes. In the 2023–2024 school year, California recorded 191 students for "Otomian (Mixteco)," 32 for Mayan languages, and 8 for Zapotec. In contrast, data gathered by Comunidades

Indígenas en Liderazgo (CIELO), a Los Angeles-based nonprofit, revealed a broader picture of language diversity. During the 2020 pandemic, CIELO distributed financial support to Indigenous communities and collected demographic and language preference data, which was used to create an Indigenous language diversity map. This map showed that Los Angeles County is home to at least 25 distinct Indigenous communities speaking 36 different languages, including Zapotec, Chinantec, K'iche', Mixe, and Q'anjob'al (Miller, 2021). Among these, the Maya family, Zapotec branch, and Purépecha were identified as the largest groups—yet this level of detail is missing from California's official state language records. Kansas, a new destination state for immigrants from Mesoamerican backgrounds and children identified as English Learners (Gómez-Cervantes, 2023; Sugarman, 2016), faces similar challenges. Despite the presence of Indigenous immigrants from Mixtec, Tlapaneco, Chuj, and K'iche' Mayan backgrounds, the current Kansas codes allow schools to identify Zapotec (zap), Mayan (myn), Central American Indian (cai), and North American Indian (nai), but omit Nahuatl (nah) and Otomian (oto). This omission leaves out individuals who may use one of the 53 Mixtec languages or any of the over 34 Mayan languages, despite reports of unaccompanied minors from Mayan backgrounds in Kansas whose first languages include K'iche' and Chuj, among others (Menjívar & Cervantes, 2024). Kansas demonstrates the need for a more nuanced approach to language documentation to accurately reflect the linguistic diversity of their student populations.

When examining the languages that appear with greater granularity in the seventeen states using ISO 639-3 codes, specific Mayan languages stand out. For instance, all states that include language-level additions from ISO 639-3 list specific Mayan languages. This means instead of using the ISO 639-2 broad Mayan language code (myn), states give their schools

specific Mayan language options such as K'iche', Chuj, Q'eqchi', Mam, and Q'anjob'al. Mayan languages are perhaps more likely to appear individually in language data because each one has a distinctive name assigned by its own language community, facilitating unique identification. This contrasts with languages like Zapotec, which are often grouped under branch names, making individual languages less likely to be separately recognized (Pérez Báez, 2013, 2017). In fact, aside from Oregon, and Colorado, there were no other states that included any Zapotec or Nahuatl languages with more specificity. Regarding specific Nahuatl languages or Mixtec languages, Minnesota, Colorado, Oregon, and Virginia either added specific languages from ISO 639-3 or added a community name. That is, states that go beyond ISO 639-3, such as Florida which includes "Chalchiteco" or Minnesota which includes "nāhuatlahtōlli" suggest that local or community-provided names were added to their lists.

To better understand how Mesoamerican languages were added to state department lists and the factors influencing their inclusion, a survey was sent to state representatives from states that reported at least five Mesoamerican languages in their language code lists. While the responses provide preliminary insights into the factors influencing modifications to state language codes, their generalizability is limited by the small sample size and varying respondent familiarity with state reporting practices.

Some states, such as Minnesota and Washington, indicated that updates to language lists were facilitated by student information systems or district requests. For example, Minnesota's transition to a three-digit language code system enabled the identification of over 440 languages. This suggests that states with adaptable systems or strong relationships with local school districts may be better positioned to meet evolving linguistic needs, though further research is needed to confirm this relationship. Respondents also pointed to collaborations among SEA leadership,

local school districts, and community partners as key drivers of these updates. Some states emphasized community-driven efforts or responsiveness to district requests as central to their processes. However, these findings should be interpreted with caution, as they may not fully capture the complexities of state-level practices or broader systemic influences on language reporting. Finally, the responses suggest that clearer federal guidance, additional training, and targeted resources could help states strengthen their linguistic documentation practices. These observations align with existing research highlighting the need for specific federal oversight to ensure consistent implementation of policies for multilingual learners (Bartlett et al., 2024; Callahan et al., 2023). Moving forward, incorporating community voices and linguistic expertise into state-level decision-making could promote more inclusive practices that better reflect the linguistic diversity of student populations. Given the survey's limitations, however, these suggestions remain exploratory and warrant further investigation.

RQ2: What is the immediate impact of introducing detailed Indigenous Mesoamerican language codes in 2020 on the number of these languages identified and recorded by the state?

Research question two provides insight into the impact of state-level actions introducing 150 new language codes that schools in Oregon could use to document students' native languages at enrollment. After the introduction of specific Mesoamerican language codes in 2020, the enrollment of individual languages within the Zapotec, Nahuatl, and Mixtec branches was documented at a greater level of granularity, which was not possible before when only relying on the ISO 639-2 codes. After the policy changed, specific languages like Chichicapan Zapotec, Guevea de Humboldt Zapotec, Isthmus-Cosoleacaque Nahuatl, Diuxi-Tilantongo Mixtec, Santa Lucía Monteverde Mixtec, and Copala Triqui were recorded and accounted for in

Oregon schools, though their enrollment numbers ranged from one to seven students. Notably, languages within the Mayan family were identified in larger numbers than previously with a reduction noted for the broad label for Mayan languages (myn) being used less frequently, decreasing from 258 in 2022 to 113 in 2023. While lower broad label Mayan was used, there was an observed increase of specific Mayan languages enrolled. For instance, Mam started with 125 students in 2020, increasing to 161 in 2021, 205 in 2022, and reaching 349 in 2023. Additionally, Chuj was first identified in 2020 with six students and is now being reported at 16 in 2023. Finally, Q'anjob'al started with two students in both 2020 and 2021, increased to 16 in 2022, and further rose to 44 in 2023. Essentially, the inclusion of these specific language codes led to an increasing trend of more specific languages being identified, especially within the Mayan language family. The regression discontinuity (RD) analysis provided further evidence on the impact of detailed language code introduction on broader language documentation practices. The results suggest that the policy change did not lead to a statistically significant reduction in the use of broad "Other" language labels as anticipated, although there was a slight decrease in this categorization. The lack of a significant change points to the persistence of broader, generalized classifications, which may hinder the recognition of the specific linguistic identities of students. This finding, combined with the enrollment data, implies that while policy changes aimed at introducing greater language granularity are a step forward, their practical impact on language documentation may be limited without additional support and implementation efforts. In fact, an Oregon SEA representative indicated that while guidance is provided to districts, it is not sufficient to accurately document Mesoamerican Indigenous languages.

Limitations

First, while acquiring language codes and searching for publicly available data, the dates of data collection were recorded. However, given the dynamic nature of websites, some links or documents may become outdated, and updates occurring after data collection may not have been captured. For instance, during the writing of the results, I revisited the links for state language codes and found that Indiana's code list link had become inactive. Upon further investigation, I discovered that the original code list, acquired on 8/12/2024, had been updated by 11/20/2024 with a new link. The updated list also included a new Mayan language, Poqomchi, and previously missing "Central American Indian Language" codes. This demonstrates that the findings of this study represent a snapshot in time and may not reflect subsequent changes and ongoing examination of language codes may be required. Future studies could investigate the timeline and adoption of federal language guidelines and determine if there are any differences to when states respond to federal guidance.

Second, the use of regression discontinuity (RD) analysis, while effective for examining policy changes, has limitations in capturing long-term effects or nuances in language code adoption. Given that the policy was implemented only recently (in 2020), the short timeframe may not fully reflect its broader impact on language identification practices or student outcomes. Longer-term analysis is needed to fully understand the effects of introducing new language codes. Moreover, temporal confounding (Shadish et al., 2022) may be present due to simultaneous events such as the COVID-19 pandemic, which could have independently influenced language recording practices, potentially delaying the uptake of Oregon's language code policy update. Additionally, there is a lack of qualitative data exploring the perspectives of key individuals impacted by these policies and their implementation such as educators, administrators, and families, regarding the language code changes or additions. Understanding

their motivations, attitudes, and challenges could provide valuable insights into the persistence and use of broad classification codes and the impact on student language support. Future studies may seek to gather this qualitative data to better understand the barriers and facilitators to use of more specific language codes by district staff.

Third, additional languages may be appearing in state code lists that are relevant to the immediate student population but are not Mesoamerican languages. Notably, some descriptive labels for languages beyond ISO 639-3 may be included in states that were not relevant to Mesoamerican languages such as Native Alaskan languages. For example, a language within the Athabascan family was included in the Alaska state code and specified as Deg Xinag, which has historically been referred to by the pejorative term "Ingalik." The use of this more appropriate label suggests that the community itself influenced the selection. Similarly, there may be other languages in other state codes that may reflect greater awareness for their student population that were not captured in this study. Future studies may examine language codes beyond Mesoamerican languages and investigations may explore the role of local communities in shaping state policy. This approach may provide a broader understanding of how state practices evolve to meet the needs of their unique student populations.

Fourth, the regression discontinuity model and findings revealed two main issues. First, the model was underpowered, with only 19.3% statistical power, well below the expected threshold of 80%. The model thus had limited capacity to detect small or moderate effects of the policy change on language enrollment trends. Additionally, using state-level data introduced limitations. While focusing on state-level actions aligned with the study's goals, this aggregated approach obscures variability that may exist across districts and schools in Oregon, limiting insights into how policies affected schools differently within districts in Oregon. To address

these limitations, future studies should analyze more granular data, such as district- or school-level trends. Disaggregating data to the school level would allow each school to serve as an observation, thereby increasing the sample size. For instance, Oregon's 1,200 public schools across 197 districts could provide thousands of observations at the school or district level, enhancing statistical power and improving model estimate precision. Lastly, advanced modeling techniques, such as hierarchical linear or mixed-effects models, should be explored to account for nested data structures and better capture regional variability. Expanding the analysis to school- or district-level data would offer deeper insights to explore if the language policy effects vary based on characteristics such as rural versus urban settings or Title I versus non-Title I schools.

Lastly, this study intended to examine all 50 states' language identification practices, but despite efforts to obtain these lists, seven states did not provide their data. Specifically, representatives from Nevada, Michigan and Kentucky did not respond to multiple inquiries. Representatives from Delaware, Vermont, West Virginia and New Hampshire responded to inquiries but did not provide their lists, citing reasons such as the absence of specific lists as they adhered to ISO 639-2, or that their state adhered to requirements for federal reporting. Future studies may seek to collaborate directly with SEAs and representatives to obtain this information. Additionally, there was a low response rate to the survey sent out to understand how states with five or more Mesoamerican languages in their code lists incorporated these changes. Of the twenty-two surveys sent, two were sent to representatives from Ohio, South Carolina, Washington, Georgia, and Florida, as well as to representatives from Rhode Island and Connecticut. The survey was also sent to email addresses associated with EL departments in Illinois, South Carolina, Virginia and Minnesota. Of the nine fully completed surveys completed,

two were from Oregon, Washington and Minnesota and one was received from Colorado, South Carolina, Connecticut. Since the surveys were sent through email and there was no additional interaction with these SEA representatives, future studies may consider including a structured interview component to gather more nuanced information not captured in the survey.

Implications for future research and application

As aforementioned, Umansky and Porter's state EL policy framework emphasizes recognizing language diversity as an asset in U.S. schools through three principles: understanding student needs, accessible high-quality instruction, and system conditions. However, the framework assumes accurate identification of non-English languages through home language surveys and bypasses how languages are accounted for and recorded in schools. As this study has demonstrated, there is immense variability in how prepared states are to account for the growing linguistic diversity of students, especially when using Mesoamerican languages which do not appear in great granularity in the federal language codes required for EDPass. While the federal government provides guidance, state education agencies have flexibility in how languages can be accounted for within their student information systems. The Oregon case study illustrates that change is possible to the accuracy of language codes in a vastly more comprehensive manner, but more work is needed for this to be done properly. To bridge these gaps, Viennet and Pont's (2017) implementation framework could be useful to improve deploying policy changes.

Work in policy implementation science specific to educational policy offers a lens through which to approach the complexities of understanding how federal guidelines get translated into practice. These findings suggest that states need more flexible student information systems, require additional training, and resources to ensure accurate and comprehensive

documentation of linguistic diversity, particularly for Mesoamerican languages. Addressing these barriers is critical to ensuring children with Indigenous backgrounds whose languages are at risk of being overlooked, are properly identified within the current language identification infrastructure.

Federal Implications

The first determinant in Viennet and Pont's framework is policy design or how a policy is developed—including its rationale and practical solutions— influences its effective implementation. Viennet and Pont (2017), drawing on Bell and Stevenson (2015) and Fullan (2015), emphasize that both the inherent characteristics of a policy and any issues left unaddressed during its development can greatly impact its implementation. In the context of federal guidance, poor implementation often results from problematic policy design, such as ESSA's requirement to use the outdated ISO 639-2 Code List. This list includes only 500 codes assigned primarily to language families or branches, severely disregarding the over 8,000 languages spoken globally (Hammarström et al., 2024). Although the federal government continues to provide assistance—such as the 2024 Technical Assistance Document recommending the use of "und" for undetermined languages and "MIS/mis" for uncoded languages—the reliance on the limited ISO 639-2 code list still presents the major challenges. Recent updates in November 2024 to the Language Codes for EDPass included a new list of accepted and unaccepted codes, and in April 2024, four ISO 639-3 languages (including Quiche, a Mayan language) were added. Yet little is known about the uptake of these changes by states. Nor is there guidance or reasoning provided for the ED's continued reliance on ISO 639-2 which even requires states to report native languages as "MIS/mis" if they are not part of ISO 639-2. This limits which languages can be accurately identified, ultimately impacting the validity of

information schools gather to make appropriate education decisions (Castro et al., 2023; Renn, 2009; Young et al., 2018). To improve implementation, the federal government should consider adopting ISO 639-3 which is the leading coding system that focuses on specific language varieties (Hammarström, 2015). Use of ISO 639-3 would provide greater insight to the individual languages a student uses in the home, enabling better recognition of languages that would otherwise be subsumed under broad, less useful labels, thereby supporting schools in understanding their students' linguistic diversity.

Action members

The second determinant in Viennet and Pont's framework is Stakeholder Engagement, which refers to the individuals and organizations responsible for implementing education policies. These stakeholders are crucial to the success of policy implementation. According to Viennet and Pont, Education policies are shaped not only by their design but also by the people involved in their execution. These "actors" include both formal entities, such as state education agencies, and informal ones, like parents and community groups.

State policies play a central role in meeting federal requirements, with SEAs responsible for overseeing educational outcomes, administering federal grants, and developing state plans (Brewer et al., 2013). Survey data from this study indicated that states incorporating more localized language labels beyond ISO 639-2 and ISO 639-3 actively engaged community members and parents in decision-making. Effective engagement of community members, including compensating them for their time, is essential for meaningful policy implementation. However, incorporating Mesoamerican languages requires ongoing training for state department staff to ensure that the information gathered from community members is appropriately incorporated. The complexity of Mesoamerican languages and the many spelling variations, as

indicated by Washington State representatives, necessitates continuous support. SEAs may need assistance from linguists and other language experts to improve understanding, accurately document languages, and avoid the use of pejorative labels or incorrect classifications. For instance, states like Virginia, Colorado, Florida, and Washington have included the term "Tarascan," which is considered derogatory, while others have mistakenly categorized Mixtec as part of the Otomian branch, despite them being independent branches (Josserand, 1983). Additionally, some states continue to use broad labels such as "Aztec," which inaccurately collapses an entire branch of 31 recognized languages under a single, non-specific category (Hammarström et al., 2024; Pérez & Vásquez, 2024; Pollard, 2012)

Another key group of stakeholders includes district-level school staff, particularly office staff involved in initial contact with families during school registration. Their ability to engage in meaningful dialogue with caregivers is crucial for effective policy implementation, as is addressing any biases they may hold toward non-English-speaking families. Prior research has shown that educators and school staff often have emerging knowledge about supporting multilingual learners, but deeper understanding is needed, particularly around Mesoamerican languages and cultures, which remains emerging in schools (Baecher et al., 2012; Baquedano-Lopez, 2019; Campbell-Montalvo, 2021; 2023). Finally, understanding parental perspectives on documentation procedures is central to improving implementation at the state level. Parental engagement with schools is typically shaped by a U.S.-centric cultural perspective, which can make parents—especially those from non-White backgrounds—feel unwelcome and disconnected (Quiocho & Daoud, 2006). It is vital for caregivers, particularly those from marginalized communities, to feel safe discussing their home language practices. This is

especially important for Mesoamerican Indigenous families, who may be reluctant to assert their Indigenous identity due to historical persecution (Moffet, 2018).

Contextual Considerations

The third determinant from Viennet and Pont's framework is contextual factors, which refer to formal and informal institutional constraints within an education system and their interaction with other existing policies in education and other sectors. These factors can either support or obstruct the implementation process. While federal mandates set broad educational frameworks, states are responsible for executing these requirements, resulting in varied implementation nationwide. State-level execution of federal laws differs immensely (Monogan, 2013; Villegas & Pompa, 2020), with past research suggesting factors such as the rate of immigration to each state can influence these differences (Callahan et al., 2023; Terrazas, 2011). For example, states like Nebraska, Kansas, and Texas are experiencing increased numbers of Mesoamerican immigrants due to agricultural opportunities. Despite this, only Nebraska has included Q'anjob'al on its state language list. Notably, the presence of Comunidad Maya Pixan Ixim, a Maya Q'anjob'al non-profit organization based in Omaha, Nebraska, advocates for Indigenous language accessibility (Comunidad Maya Pixsn Ixim, 2024). Operating across legal and educational sectors, securing over $2.2 million in contributions and grants in 2023, the organization promotes the teaching of Q'anjob'al through its Maya Education initiative. Their efforts may provide insights into how Q'anjob'al came to be added to Nebraska's code list. It is unclear if Texas or Kansas has similar community groups to assist in advocacy.

Notably, previous research shows significant variation in how states respond to immigrant and language policies (Callahan et al., 2023; Lee & Hawkins, 2015). Therefore, it is crucial to consider how demographic changes—particularly the presence or absence of new

immigrant populations—may influence the need for updated language codes. In this study, we observed considerable differences in how native languages are identified across states and how states set their definitions for non-English languages that are present to a significant extent (ESSA, 2015). However, it is still unclear how these variations impact student outcomes.

Implementation Strategies

The final determinant is Implementation Strategy defined as a well-defined operational plan that is essential for guiding the execution of policy (Viennet & Pont, 2016). Key elements of the implementation plan include task allocation and accountability, objectives and tools, resources, timing, and strategies for effective communication and engagement with stakeholders.

One immediate step to improve the inclusion of languages within state language codes is to adopt a centralized student information system. These systems store critical student-level data, including language information (Kusakunniran et al., 2014). Notably, a state representative from Minnesota indicated that switching to a new student information system allowed them to use a three-digit language code system, increasing the number of identifiable languages in their state's language list. Upon closer examination, I confirmed that The Minnesota Automated Reporting Student System (MARSS) is a centralized system that collects data for the Minnesota Department of Education from all Local Education Agencies, meaning there are not multiple systems used by different districts, which may improve the consistency and accuracy of data reporting. An alternative to using a centralized student information system is allowing districts to select their own systems. In Oregon, as reported by a program specialist, multiple student information systems are in use. Oregon may benefit from consolidating to a single system, as a

centralized student information system would ensure that data such as language data is uniform as well as making professional development in its use easier for the SEA.

The survey results stress the need for additional guidance and training for SEAs. A needs-based assessment (Rossi et al., 2019) is required to determine existing knowledge about language codes, student information systems, and Mesoamerican languages. Training could be led by teachers and language specialists—including speech-language pathologists and English language specialists—to develop resources and professional development opportunities that address the linguistic diversity specific to each state. This effort may also involve collaborations with experts from disciplines beyond education, such as linguistics. Central to this effort is dispelling persistent myths about Mesoamerican languages, such as the misconceptions that they are extinct, that they are merely dialects of Spanish, or that languages named after branches (e.g., "Mixtec") represent a singular, homogenous language (Pérez Báez, 2013, 2017). This is critical information that state department representatives need to understand that when an interpreter is required for school meetings, assessment or to communicate with a caregiver, requesting a "Mixtec" speaker without recognizing the significant linguistic differences within Mixtec languages could result in inequitable access to information for caregivers and quality supports to students (Araujo, 2009; Baquedano-López et al., 2013; Campbell-Montalvo, 2021).

In the Oregon case study, they indicated that training was provided to some schools when the language codes were updated, but their impact and how to expand these efforts remain unclear. Accurate documentation of languages in schools is not just a technical task or a procedure; it is crucial for improving educational outcomes for Mesoamerican students and ensuring their cultural and linguistic identities are recognized and valued. This work is an initial step toward addressing systemic inequities and fostering a more inclusive environment for all

multilingual learners. Effective training that includes cultural competency and language documentation skills can help to address systemic inequities. Many educators may lack familiarity with Mesoamerican Indigenous languages and their continued existence, which can result in the marginalization of these students if not acknowledged much less inappropriate language supports for these students. Professional development should not only focus on technical aspects of language identification but also emphasize the importance of building trust with families from marginalized backgrounds, understanding historical contexts of Indigenous communities, and fostering an inclusive classroom culture (Kiramba et al., 2022).

A Call to Action

The following recommendations are drawn from key findings of this study. To account for the growing linguistic diversity in schools—particularly Mesoamerican languages—we must move beyond broad labels and outdated systems. These steps provide a clear, actionable path for federal, state, and local education leaders to ensure that every school is better prepared recognize Mesoamerican languages in their schools.

Federal Policy Improvements

- Adopt ISO 639-3 and replace the outdated ISO 639-2 as the source list for language codes states can use to create their state code lists. ISO 639-3 provides greater specificity and includes a broader range of global languages.
- Improve guidance and technical assistance provided to states. The U.S. Department of Education (ED) should establish clear implementation timelines and accountability measures to ensure states consistently update their language codes to accurately reflect their student populations.
- ***State Level Actions***

- Adopt centralized student information systems to improve the consistency, accuracy of language data reporting sent to the ED.
- Engage and actively collaborate with community groups, parents, language experts such as linguist to ensure accurate language labels are used in their code lists.
- Conduct a needs-based assessment to evaluate state staff's knowledge about language codes, student information system and the language needs of their code lists.

District and School Level Actions

- Support Registration and initial contact by training office staff to meaningfully engage with families during initial school registration to ensure accurate language identification and documentation is taking place.
- Provide culturally responsivity training. One topic could be training on Mesoamerican languages and culture with an emphasis on their diversity. Dispelling myths (that they are extinct or dialects of Spanish).

Conclusion

As the population of diverse students continues to grow, it is essential that schools possess accurate information about the languages of their students (Kiramba et al., 2022). The findings from this study illustrate that accurate language identification and recording requires not only a well-defined policy but also sufficient training and resources for those on the frontlines of implementation. This study calls for a coordinated effort among SEAs, educators, and communities to ensure that language documentation practices are inclusive, equitable, and reflective of the rich linguistic diversity present in U.S. schools. By integrating Viennet and Pont's implementation-focused framework with Umansky and Porter's framework, action-oriented implications for practice become apparent.

References

Abedi, J. (2008). Classification System for English language learners: Issues and Recommendations. *National Council on Measurement in Education*, 17-31.

Acuña, Rudy. 2007. Corridors of Migration: The Odyssey of Mexican Laborers, 1600–1933. Tucson: University of Arizona Press.

Araujo, B. E. (2009). Best practices in working with linguistically diverse families. Intervention in School and Clinic, 45(2), 116-123. doi:https://doi.org/10.1177/1053451209340221

Artiles, A., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. Exceptional Children, 71, 283–300. doi:10.1177/001440290507100305

Artiles, A. J., Kozleski, E. B., Trent, S. C., Osher, D., & Ortiz, A. (2010). Justifying and explaining disproportionality, 1968–2008: A critique of underlying views of culture. *Exceptional Children*, *76*(3), 279–299. https://doi.org/10.1177/001440291007600303

Ahtone, T., & Ahtone, T. (2018). Indigenous immigrants face unique challenges at the border. High Country News. https://www.hcn.org/articles/tribal-affairs-indigenous-immigrants-face-unique-challenges-at-the-border/

Bailey, A. L., & Kelly, K. R. (2012). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy*, 27(5), 770–804. https://doi.org/10.1177/0895904811432137

Bailey, A. L., & Carroll, P. E. (2015). Assessment of English Language learners in the era of new academic Content Standards. *Review of Research in Education*, 39(1), 253–294. https://doi.org/10.3102/0091732x14556074

Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. Canadian Journal of Statistics, 27, 3-23.

Baquedano-López, P., Alexander, R. A., & Hernandez, S. J. (2013). Equity issues in parental and community involvement in schools: What teacher educators need to know. Review of Research in Education, 37(1), 149-182.

Baquedano-López, P. (2019). CHAPTER 9 Indigenous Maya Families from Yucatán in San Francisco: Hemispheric Mobility and Pedagogies of Diaspora. In Accountability across borders: Migrant rights in North America (pp. 239-259). University of Texas Press.

Batz, G. (2014). Maya Cultural Resistance in Los Angeles The Recovery of Identity and Culture among Maya Youth. Latin American Perspectives, 41, 194–207. https://doi.org/10.1177/0094582X14531727

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. Bilingualism, 15(3), 616–629. https://doi.org/10.1017/s1366728912000090

Bilingual Education Act [BEA] of 1968 Elementary and Secondary Act Amendments, Pub. L. No. 90–247, title VII, §702, 81 Stat. 816. (1968).

Bishop L., Kelley P. (2013). Indigenous Mexican languages and the politics of language shift in the United States. In Benson C., Kosonen K. (Eds.), Language issues in comparative education (pp. 97–113). Sense. https://doi.org/10.1007/978-94-6209-218-1_6

Brooks, M. D. (2018). Pushing past myths: Designing instruction for long-term English learners. *TESOL Quarterly*, 52 (1), 221–233. https://doi.org/10.1002/tesq.435 .

Burla, L., Knierim, B., Barth, J., Liewald, K., Duetz, M., & Abel, T. (2008). From text to codings - Intercoder reliability assessment in qualitative content analysis. Nursing Research (New York), 57(2), 113–117. https://doi.org/10.1097/01.NNR.0000313482.33917.7d

Callahan, R. 2013. The english learner dropout dilemma: Multiple risks and multiple resources California dropout research. http://cdrpsb.org/researchreport19.pdf.

Callahan, R., Gautsch, L., Hopkins, M., & Carmen Unda, M. D. (2023). Equity and state immigrant inclusivity: English learner education in ESSA. *Educational Policy (Los Altos, Calif.)*, *36*(5), 1011–1053. https://doi.org/10.1177/0895904820925819

Campbell, L., & Kaufman, T. (1983). Mesoamerican Historical Linguistics and Distant Genetic Relationship: Getting It Straight. American Anthropologist, 85(2), 362–372. https://doi.org/10.1525/aa.1983.85.2.02a00080

Campbell, L., Kaufman, T., Smith-Stark, T. C. (1986). Meso-America as a linguistic area. Linguistic Society of America, 62(3). 530-570. https://doi.org/10.2307/415477

Campbell, E. W. (2017). Otomanguean historical linguistics: Past, present, and prospects for the future. *Language and Linguistics Compass*, 11(4). Portico. https://doi.org/10.1111/lnc3.12240

Campbell-Montalvo, R. (2021). Linguistic Re-Formation in Florida Heartland Schools: School erasures of Indigenous Latino Languages. *American Educational Research Journal*, *58*(1), 32–67. https://doi.org/10.3102/0002831220924353

Campbell-Montalvo, R. A. (2023). The Latinization of indigenous students: Erasing identity and restricting opportunity at school. Lexington Books, an imprint of The Rowman & Littlefield Publishing Group, Inc.

Canty, Angelo; Ripley, B. D. (2024). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-31. https://cran.r-project.org/web/packages/boot/citation.html

Castañeda v. Pickard, 648 F.2d 989

Castilla-Earls, A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Pena, E. (2020). Beyond Scores: Using Converging Evidence to Determine Speech and Language Services Eligibility for Dual Language Learners. *American Journal of Speech-Language Pathology*, *29*(3), 1116–1132. https://doi.org/10.1044/2020_AJSLP-19-00179

Castilla-Earls, A., Francis, D., Iglesias, A., & Davidson, K. (2019). The Impact of the Spanish-to-English Proficiency Shift on the Grammaticality of English Learners. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1739–1754. https://doi.org/10.1044/2018_JSLHR-L-18-0324

Castro, A., Siegel-Hawley, G., Bridges, K., & Williams, S. E. (2023). Drawn Into Policy: A Systematic Review of School Rezoning Rationales, Processes, and Outcomes. Review of Educational Research, 0(0). https://doi-org.uoregon.idm.oclc.org/10.3102/00346543231195816

Champely, Stephane; Ekstrom, Claus; Dalgaard, Peter; Gill, Jeffrey; Weibelzahl, Stephan; Anandkumar, Aditya; Ford, Clay; Volcic, Robert; De Rosario, Helios. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

Cohen, D. (2011). Braceros migrant citizens and transnational subjects in the postwar United States and Mexico. University of North Carolina Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Conger, A. J. (2017). Kappa and rater accuracy: Paradigms and parameters. Educational and Psychological Measurement, 77, 1019-1047.

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. https://doi.org/10.1177/003172171109300316

Cooc, N. (2023). National trends in special education and academic outcomes for English learners with disabilities. *The Journal of Special Education*, 57(2), 106–117. https://doi.org/10.1177/00224669221147272

Cruz, R., & Rodl, J. (2018). An integrative synthesis of literature on disproportionality in special education. *The Journal of Special Education*, 52(1), 50–63. https://doi.org/10.1177/0022466918758707

Cuiccio, C., & Husby-Slater, M. (2018). Needs Assessment Guidebook: Supporting the Development of District and School Needs Assessments. State Support Network.

Davis, J. W., & Bauman, K. J. (2008). School Enrollment in the United States: 2006. Population Characteristics. Current Population Reports. US Census Bureau.

DeCamp, J. (2008). A Case Study Related to Standardization–Codes for the Representation of Names of Languages. Uses and usage of language resource-related standards, 9.

DiCanio, C. T., Zhang, C., Whalen, D. H., & García, R. C. (2019). Phonetic structure in Yoloxóchitl Mixtec consonants. *Journal of the International Phonetic Association*, *50*(3), 333–365. https://doi.org/10.1017/s0025100318000294

Diem, S., Young, M. D., Welton, A. D., Mansfield, K. C., & Lee, P. L. (2014). The intellectual landscape of critical policy analysis. International Journal of Qualitative Studies in Education, 27(9), 1068–1090. https://doi.org/10.1080/09518398.2014.916007

Drude S. Reflections on diversity linguistics: Language inventories and atlases. In: McDonnell B, Berez-Kroeker AL, Holton G, editors. Reflections on language documentation 20 years after Himmelmann 1998. 2018. p. 122–31.

Dunn, M. H. (2000). Privatization, land reform, and property rights: the Mexican experience. *Constitutional Political Economy*, 11, 215-230.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015).

Estrada, P. (2014). English learner curricular streams in four middle schools: Triage in the trenches. *The Urban Review*, 46(5), 535.11(36), 1–54

Equal Education Opportunities Act, Pub. L. 93-380, § 204 (1974). https://www.congress.gov/bill/93rd-congress/house-bill/40

Fernández-Kelly, P., Harvey, D., Centeno, M. A., Portes, A., Massey, D. S., Wise, R. D., Cypher, J. M., Agarwala, R., Yashar, D. J., Shefner, J., Stallings, B., Wherry, F., Light, D. W., & Gellert, P. K. (2007). NAFTA and beyond: Alternative perspectives in the study of global trade and development. *The Annals of the American Academy of Political and Social Science*, 610(1), 6–19. https://doi.org/10.1177/0002716206297972

Fox, J. A., & Rivera-Salgado, G. (2004). "Introduction", Indígenas mexicanos migrantes en los Estados Unidos. Center for Global, International and Regional Studies. https://escholarship.org/content/qt6pb2j4bt/qt6pb2j4bt.pdf?t=lnr4ko

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.

Gamer M; Lemon J; Fellows, Ian; Puspendra, Singh. (2019). *irr: Various coefficients of interrater reliability and agreement*. R package version 0.84.1, https://CRAN.R-project.org/package=irr.

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. https://doi.org/10.14507/epaa.v11n36.2003

Gándara, P. (2015). Charting the relationship of English learners and the ESEA: One step forward, two steps back. *The Russell Sage Foundation Journal of the Social Sciences*, *1*(3), 112. https://doi.org/10.7758/rsf.2015.1.3.06

Gómez Cervantes, A. (2023). Language, race, and illegality: indigenous migrants navigating the immigration regime in a new destination. *Journal of Ethnic and Migration Studies*, *49*(7), 1610–1629. https://doi.org/10.1080/1369183X.2021.1968294

Goodrich, Fitton, L., Chan, J., & Davis, C. J. (2023). Assessing oral language when screening multilingual children for learning disabilities in reading. Intervention in School and Clinic, 58(3), 164–172. https://doi.org/10.1177/10534512221081264

Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. Child Indicators Research, 7(4), 821–841. https://doi.org/10.1007/s12187-013-9230-6

Guiberson, M., & Atkins, J. (2012). Speech-language pathologists' preparation, practices, and perspectives on serving culturally and linguistically diverse children. *Communication Disorders Quarterly*, 33(3), 169–180. https://doi.org/10.1177/1525740110384132

Gutiérrez, K. D., & Orellana, M. F. (2006). At last: The "problem" of English learners: constructing genres of difference. *Research in the Teaching of English*, 40(4), 502–507. https://doi.org/10.58680/rte20065110

Hammarström H., Forkel, R., Haspelmath, M., & Bank, S. (2023). *Glottolog (4.8)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://glottolog.org

Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2024. Glottolog 5.0. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.10804357

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*(4), 715–733. https://doi.org/10.1016/j.ecresq.2014.05.008

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E. (2018). Bilingual Development in Children of Immigrant Families. *Child Development Perspectives*, 12(2), 80–86. https://doi.org/10.1111/cdep.12262

Holmes, S. M. (2013). "Is it worth risking your life?": Ethnography, risk and death on the US–Mexico border. Social Science & Medicine, 99, 153-161.

Huizar Murillo, J., & Cerda, I. (2004). Indigenous Mexican migrants in the 2000 US Census:"Hispanic American Indians.". Indigenous Mexican Migrants in the United States, 279-302.

Individuals With Disabilities Education Act, 20 U.S.C. §§ 1400 et seq. (2006 & Supp. V. 2011)

Josserand, J. (1983). Mixtec dialect history.

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. https://doi.org/10.3102/0002831220915702

Kiramba, L. K., Deng, Q., & Viesca, K. (2022). Novice general education teachers' perceptions of preparedness in U.S. public schools: The impact of learning about and working with multilingual students. *Teaching and Teacher Education*, *116*, 103757. https://doi.org/10.1016/j.tate.2022.103757

Krogstad, J. M., Passel, J. S. & Cohn, D. (2019). 5 facts about illegal immigration in the U.S. Pew Research Center. Retrieved from: https://www.pewresearch.org/short-reads/2019/06/12/5-facts-about-illegal-immigration-in-the-u-s/

Lacelle-Peterson, M., & Rivera, C. (1994). Is it real for all kids? A framework for Equitable assessment policies for English language learners. *Harvard Educational Review*, *64*(1), 55–76. https://doi.org/10.17763/haer.64.1.k3387733755817j7

Lau v. Nichols, 414 U.S. 563 (1964).

Library of Congress (2017). *ISO 639-2 Language Code List - Codes for the representation of names of languages*. Retrieved from https://www.loc.gov/standards/iso639-2/php/code_list.php

Linquanti, R., Cook, H. G., & Council of Chief State School Officers. (2013). Toward a "Common Definition of English Learner" a brief defining policy and technical issues and opportunities for state assessment consortia. https://files.eric.ed.gov/fulltext/ED542705.pdf

Linquanti R., Cook H. G. (2015). Re-examining reclassification: Guidance from a national working session on policies and practices for exiting students from English learner status. Washington, DC: Council of Chief State School Officers.

Llahmon, C. E. & Gupta, V. (2015). Dear colleague letter: English learner students and limited English proficient parents. United States Department of Justice & United States Department of Education of Office of Civil Rights. https://www2.ed.gov/about/offices/list/ocr/letters/colleague-el-201501.pdf

López Austin, A., & López Luján, L. (2001). Mexico's indigenous past (English-language edition.). University of Oklahoma Press.

López, F., & Iribarren, J. (2014). Creating and Sustaining Inclusive Instructional Settings for English Language Learners: Why, What, and How. *Theory into Practice*, 53(2), 106–114. https://doi.org/10.1080/00405841.2014.885810

Loza, Mireya. (2016) "Yo Era Indígena: Race, modernity, and the transformational politics of transnational labor." Defiant Braceros: How Migrant Workers Fought for Racial, Sexual, and Political Freedom, University of North Carolina Press, pp. 23–62.

Machado-Casas, M. (2012). Pedagogías del Camaleón /Pedagogies of the Chameleon: Identity and Strategies of Survival for Transnational Indigenous Latino Immigrants in the US South. *Urban Review*, 44(5), 534–550. https://doi.org/10.1007/s11256-012-0206-5

Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish speaking language minority learners. *Journal of Educational Psychology*, 102(3), 701–711. https://doi.org/10.1037/a0019135

Martínez, R. A., & Mesinas, M. (2019). Linguistic motherwork in the Zapotec diaspora: "Zapoteca" mothers' perspectives on indigenous language maintenance. Association of

Mexican American Educators Journal, 13(2), 122-. https://doi.org/10.24974/amae.13.2.431

Martínez, R. A., & Martinez, D. C. (2022). Learning in dialogue with Latinx children of immigrants: reflections on the co-emergence of collaborative linguistic inquiry and critical pedagogical praxis. *Urban Education*, 4208592210826-https://doi.org/10.1177/00420859221082670

Massey, D. S., & Sana, M. (2003). Patterns of US migration from Mexico, the Caribbean, and Central America. Migraciones internacionales, 2(2), 05-39.

Matheny, K. T., Thompson, M. E., Townley-Flores, C., & Reardon, S. F. (2023). Uneven progress: Recent trends in academic performance among US school districts. *American Educational Research Journal*, 60(3), 447-485.

Medina J. (2019). Anyone Speak K'iche' or mam? Immigration courts overwhelmed by indigenous languages. The New York Times. https://www.nytimes.com/2019/03/19/us/translators-border-wall-immigration.html

Menjívar, C., & Gómez Cervantes, A. (2024). Maya Guatemalans Seeking Asylum: Race and Gender in a Continuum of State Control. *Sociology of Race and Ethnicity*, *0*(0). https://doi-org.uoregon.idm.oclc.org/10.1177/23326492241238945

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. doi:10.1080/19313152.2012.665822

Miksch, K. (2019). Civil Rights Act of 1964. Education Law. https://usedulaw.com/216- civil-rights-act-of-1964.html

Miller, L. (2021, July 7). New map highlights L.A.'s Indigenous communities - Los Angeles Times. *Los Angeles Times*. https://www.latimes.com/california/story/2021-07-07/la-me-indigenous-map-los-angeles

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

National Center for Education Statistics. (2023, May). English Language learners in public schools. Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/programs/coe/indicator/cgf/english-learners

National Center for Education Statistics. (2024). English Learners in Public Schools. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/programs/coe/indicator/cgf.

Nolan, R. (2019, December 30). A translation crisis at the border. The New Yorker. https://www.newyorker.com/magazine/2020/01/06/a-translation-crisis-at-the-border

Obinna, D. N. (2023). "Alone in a crowd: Indigenous migrants and language barriers in American immigration". Race and Justice, 13(4), 488-505. https://doi.org/10.1177/21533687211006448

Office of English Language Acquisition. (2022). *English learners: Demographic trends*. United States Department of Education. /https://ncela.ed.gov/sites/default/files/2022-09/ELDemographics_20220805_508.pdf

Office of Elementary and Secondary Education. (2024). Office of Elementary and Secondary Education. https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/english-language-acquisition-state-grants/performance/

Oetting, J. B. (2018). Prologue: Toward accurate identification of developmental language disorder within linguistically diverse schools. *Language, Speech & Hearing Services in Schools*, 49(2), 213–217. https://doi.org/10.1044/2018_lshss-clsld-17-0156

Park, M., Zong, J., & Batalova, J. (2018). Growing superdiversity among young US dual language learners and its implications. Washington, DC: Migration Policy Institute.

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, 50(4), 793–817. https://doi.org/10.1017/S0305000922000708

Peña, E. D., Bedore, L. M., & Torres, J. (2021). Assessment of language proficiency and dominance in monolinguals and bilinguals. In Bilingualism across the lifespan (pp. 88-105). Routledge

Peñalosa. (1985). Trilingualism in the Barrio: Mayan Indians in Los Angeles. Language Problems and Language Planning, 10(3), 229–252.

Pérez-Báez, Gabriela. 2013. "Family language policy, transnationalism, and the diaspora community of San Lucas Quiaviní of Oaxaca, Mexico." *Language Policy*. 12 (1):27–45. https://doi.org/10.1007/s10993-012-9270-7

Pérez Báez, G., Vidaurri, C., & Barreiro, J. (2022). Indigenous Latino populations in the United States. In I. Krupnik (Ed.) Handbook of North American Indians, Vol. 1. Smithsonian Scholarly Press. https://doi.org/10.5479/si.21262173.

Pérez, W., & Vásquez, R. (2024). *Culturally responsive schooling for indigenous Mexican students*. Multilingual Matters.

Plyler v. Doe, 457 U.S. 202 457 U.S. 202, No. 80-1538. 1982.

Pollard, H.P. (2012) The Tarascan empire: Postclassic social complexity in western Mexico. In D.L. Nichols and C.A. Pool (eds) The Oxford Handbook of Mesoamerican Archaeology (pp. 434–448). Oxford University Press

Pompa, D., & Villegas, L. (2017). Analyzing state ESSA plans for English learner accountability: A framework for community stakeholders. Washington, DC: Migration Policy Institute.

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

Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Pulido, Laura, Laura Barraclough, and Wendy Cheng. 2012. A People's Guide to Los Angeles. Berkeley: University of California Press.

Renn, K. A. (2009). Educational policy, politics, and mixed heritage students in the United States. *Journal of Social Issues*, 65(1), 165–183. https://doi.org/10.1111/j.1540-4560.2008.01593.x

Reynolds, J. F. (2013). (Be)laboring Childhoods in Postville, Iowa. *Anthropological Quarterly*, *86*(3), 851–889. http://www.jstor.org/stable/43652639

Riva, S. (2022). Tracing invisibility as a Colonial Project: indigenous women who seek asylum at the US-Mexico Border. Journal of immigrant & refugee studies, 20(4), 584-597.

Rossi, P. H. (Peter H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: a systematic approach* (Eighth edition.). SAGE.

Ruiz, N. T., & Barajas, M. (2012). Multiple Perspectives on the Schooling of Mexican Indigenous Students in the U.S.: Issues for Future Research. Bilingual Research Journal, 35(2), 125–144. https://doi.org/10.1080/15235882.2012.703639

Salerno, A. S., & Andrei, E. (2021). Inconsistencies in English Learner identification: An inventory of how home language surveys across U.S. States screen multilingual students. AERA Open, 7(1), 1-16. https://doi.org.uoregon.idm.oclc.org/10.1177/23328584211002212

Sandoval, G. F. (2014). Immigrant integration models in "illegal" communities: Postville Iowa's shadow context. *Local Environment*, *20*(6), 683–705. https://doi-org.uoregon.idm.oclc.org/10.1080/13549839.2014.963839

Sanford, Victoria. 2003. Buried Secrets: Truth and Human Rights in Guatemala. New York: Palgrave Macmillan.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (Donald T. (2001). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

SIL International *ISO 639-3* |. (2024). https://iso639-3.sil.org/

Simons, G. (2013, February 9). ISO 639-3: Where are we and how did we get here? Presented at the Workshop on Identifying Codes for Languages, Newcastle, Australia. Retrieved from https://users.monash.edu.au/~smusgrav/ARGILaRe/Workshop/Simons.pdf

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

Stephen, Lynn. 2007. Transborder Lives: Indigenous Oaxacans in Mexico, California, and Oregon. Durham, NC: Duke University Press.

Stephen, L., & Velasco-Ortiz, L. (2023). Introduction: Mesoamerican indigenous mobilities in Mexico and the United States. Mexican Studies, 39(1), 7–31. https://doi.org/10.1525/msem.2023.39.1.7

Stephen, L. (2023). Vulnerabilities and Collective Care: Indigenous Guatemalan and Mexican Farmworkers in Diaspora Confronting COVID-19 in the Western United States. Mexican Studies, 39(1), 117–144. https://doi.org/10.1525/msem.2023.39.1.117

Suárez, J. A. (1983). The Mesoamerican Indian languages. Cambridge University Press.

Sugarman, J. (2016). Funding an Equitable Education for English Learners in the United States. In *Migration Policy Institute*. Migration Policy Institute.

Thompson, K. D. (2015). English learners' time to reclassification. *Educational Policy*, 31(3), 330–363. https://doi.org/10.1177/0895904815598394

Thompson, K. D. (2015b). Questioning the Long-Term English Learner Label: How categorization can blind us to students' abilities. *Teachers College Record*, *117*(12), 1–50. https://doi.org/10.1177/016146811511701203

Thompson, K. D., & Rodriguez-Mojica, C. (2022). Individualized Language Plans: A Potential Tool for Collaboration to Support Multilingual Students. *Journal of Education for Students Placed at Risk*, 28(1), 97–121. https://doi.org/10.1080/10824669.2022.2123330

Track Immigration (2021) *40 Languages Spoken Among Asylum Seekers with Pending MPP Cases.* https://trac.syr.edu/immigration/reports/644/#:~:text=For%20this%20report%2C%20TRAC%20examined,(1.7%20percent)%20spoke%20Portuguese.

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. https://doi.org/10.3102/0002831214545110

Umansky, I. M. (2016). Leveled and exclusionary tracking. American Educational Research Journal, 53(6), 1792–1833. https://doi.org/10.3102/0002831216675404

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. https://doi.org/10.1177/0014402917707470

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). 1-40. https://doi.org/10.14507/epaa.28.4594

US Census Bureau. (2023). *A look at the largest American Indian and Alaska native tribes and villages in the nation, tribal areas and states*. Census.gov. https://www.census.gov/library/stories/2023/10/2020-census-dhc-a-aian-population.html#:~:text=Aztec%20and%20Maya%2C%20which%20were,combination%20population%20(Table%205).

U.S. Department of Education. (2016). Non-regulatory guidance: English learners and Title III of the Elementary and Secondary Education Act (ESEA), as amended by Every Student Succeeds Act (ESSA). Retrieved from https:// www2.ed.gov/policy/elsec/leg/essa/essatitleiiiguidenglishlearners92016.pdf

U.S. Department of Homeland Security (2020). Fiscal Year 2020 Enforcement Lifecycle Report. Office of Immigration Statistics. Retrieved February 23, 2024 from https://www.dhs.gov/sites/default/files/publications/immigration-

statistics/Special_Reports/Enforcement_Lifecycle/2020_enforcement_lifecycle_report.pdf.

Velasco Ortiz, Laura. 2014. “Transnational Ethnic Processes: Indigenous Mexican Migration to the United States.” Latin American Perspectives 41 (3): 54–75.

Velasco, P. (2010). Indigenous students in bilingual Spanish–English classrooms in New York: a teacher’s mediation strategies. *International Journal of the Sociology of Language*, *2010*(206). https://doi.org/10.1515/ijsl.2010.057

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

Wiley, T. G., & Wright, W. E. (2004). Against the Undertow: Language-Minority Education Policy and Politics in the “Age of Accountability”. *Educational Policy*, 18(1), 142-168. https://doi.org/10.1177/0895904803260030

Yescas, C. (2010). Hidden in Plain Sight: Indigenous Migrants, Their Movements, and. migrationpolicy.org. https://www.migrationpolicy.org/article/hidden-plain-sight-indigenous-migrants-their-movements-andtheir-challenges

Young, C., McNamara, G., Brown, M., & O’Hara, J. (2018). Adopting and adapting: School leaders in the age of data-informed decision making. *Educational Assessment, Evaluation and Accountability*, 30(2), 133-158.

Zong, J., & Batalova, J. (2015). The limited English proficient population in the United States. Migration Policy Institute. http://www.migrationpolicy.org/article/limited-english-proficient-population-united-states#Age,%20Race,%20and%20Ethnicity

Zyskind, K., Dorman, M., Medina, Y., & Báez, G. P. (2024). Visibility for Indigenous Students and Their Languages: Analysis of Home Language Data in Federal Reports across Seven U.S. States. *Social Sciences*, *13*(8), 427. https://doi.org/10.3390/socsci13080427

Table 1

Initial IOA

Category	Subjects	Kappa	Agreement Level	z-value	p-value
Full Adherence (ISO 639-2)	12	1.00	Perfect	3.46	0.000532
Partial Adherence (ISO 639-2)	12	1.00	Perfect	3.46	0.000532
No Adherence (ISO 639-2)	12	1.00	Perfect	3.46	0.000532
Inclusion of Languages Not in ISO 639-2	12	0.75	Substantial	2.68	0.00729
Typology Rating	12	0.87	Almost Perfect	4.21	2.57e-05
General Mesoamerican Classification	12	0.122	Slight	0.883	0.377
Specific Mesoamerican IDs	12	1.00	Perfect	3.46	0.000532

Table 2

Cohen's Kappa Statistics for Inter-Rater Reliability Final Typology

Category	States (N)	Kappa	Standard Error	p-value
Adding Codes Beyond ISO 639-2	13	0.85	0.16	0.002
Uses Labels from ISO 639-3	13	1.00	0.00	0.003
Beyond 639-3	13	1.00	0.00	0.003
Uses 9999/Unknown/Unlisted	13	0.69	0.21	0.0128

Table 3

Descriptive Statistics for Specific Mesoamerican Languages Added to State Codes

Statistic	Value
Mean	6.49
Median	1.00
SD	22.95
Min	0.00
Max	150.00
Range	150.00
N	43

Note. Refers to number of specific Mesoamerican languages found across the 43 state codes examined.

Table 4

Specific Mesoamerican Languages Count by State

State	Specific Mesoamerican Languages
Alabama	0
Alaska	0
Arizona	0
Arkansas	2
California	1
Colorado	19
Connecticut	6
Florida	10
Georgia	5
Idaho	0
Illinois	9
Indiana	12
Iowa	0
Kansas	0
Louisiana	0
Maine	0
Maryland	0
Massachusetts	0
Minnesota	6
Mississippi	0
Missouri	0
Montana	4
Nebraska	1
New Jersey	0
New Mexico	1
New York	0
North Carolina	7
North Dakota	0
Ohio	5
Oklahoma	0
Oregon	150
Pennsylvania	0
Rhode Island	1
South Carolina	6
South Dakota	0
Tennessee	1
Texas	0
Utah	1
Virginia	20
Washington	11

Wisconsin	0
Wyoming	0

Table 5

Survey Respondent Overview

State	Representative Title
Colorado	Director
Washington	Associate Director
Washington	Assistant Director
Oregon	Director
Oregon	Program Specialist
Connecticut	Director
Minnesota	Program Specialist
Minnesota	Program Specialist
South Carolina	Program Specialist

Table 6

Median Enrollment and English Learners by Year

Year	Median Enrolled	IQR Enrolled	Median EL	IQR EL
2016	6	59.25	7.0	43.00
2017	6	68.50	8.0	48.00
2018	8	85.00	9.0	44.00
2019	5	37.50	6.5	38.25
2020	5	39.00	5.0	39.75
2021	5	34.75	5.0	39.75
2022	5	27.50	4.0	27.00
2023	4	27.00	4.0	29.50

Note. IQR = Interquartile Range.

Table 7

Broad ISO 639-2 Labels within EL students 2016-2023

Year	Language Label	Count
2016	Mayan languages	127
2016	North American Indian	45
2016	Zapotec	6
2017	Mayan languages	173
2017	North American Indian	47
2017	Zapotec	2
2018	Central American Indian	2
2018	Mayan languages	243
2018	Nahuatl	1
2018	North American Indian	39
2018	Zapotec	0
2019	Central American Indian	2
2019	Mayan languages	306
2019	Nahuatl	2
2019	North American Indian	37
2019	Zapotec	1
2020	Central American Indian	2
2020	Mayan languages	289
2020	Nahuatl	1
2020	North American Indian	41
2020	Zapotec	1
2021	Central American Indian	2
2021	Mayan languages	289
2021	Nahuatl	1
2021	North American Indian	41
2021	Zapotec	1
2022	Central American Indian	2
2022	Mayan languages	271
2022	Nahuatl	2
2022	North American Indian	35
2022	Zapotec	1
2023	Central American Indian	1
2023	Mayan languages	110
2023	Nahuatl	6
2023	North American Indian	38
2023	Zapotec	3

Table 8

Enrollment Counts for Specific Zapotec Languages (2020–2023)

Year	Language	Total Enrolled
2021	Zapotec, Cajonos	1
2021	Zapotec, Lapaguía-Guivini	1
2022	Zapotec, Cajonos	1
2022	Zapotec, Santa Catarina Albarradas	1
2022	Zapotec, Southern Rincon	1
2022	Zapotec, Yatee	1
2023	Zapotec, Cajonos	1
2023	Zapotec, Chichicapan	1
2023	Zapotec, Guevea De Humboldt	1
2023	Zapotec, San Vicente Coatlán	1
2023	Zapotec, Santiago Xanica	1
2023	Zapotec, Southern Rincon	1

Table 9

Post Policy Enrollment Counts for Specific Nahuatl Languages (2020–2023)

Year	Language	Total Enrolled
2021	Nahuatl, Orizaba	1
2021	Nahuatl, Isthmus-Cosoleacaque	1
2022	Nahuatl, Isthmus-Cosoleacaque	1
2022	Nahuatl, Isthmus-Pajapan	1
2023	Nahuatl, Isthmus-Pajapan	0
2023	Nahuatl, Isthmus-Cosoleacaque	1
2023	Nahuatl, Classical	1

Table 10

Post Policy Specific Enrollment Mayan Label Use

Year	Language	Total Enrolled
2020	K'iche'	1
2020	Mam	125
2020	Q'anjob'al	2
2020	Chuj	1
2020	Popti'	0
2021	Mam	161
2021	K'iche'	3
2021	Aguacateco	2
2021	Chuj	6
2021	Q'anjob'al	2
2021	Popti'	1
2022	Cakchiquel, Mam	1
2022	K'iche'	6
2022	Q'anjob'al	16
2022	Aguacateco	4
2022	Mam	205
2022	Popti'	3
2022	Chuj	15
2023	Popti'	3
2023	K'iche'	16
2023	Chuj	24
2023	Cakchiquel, Mam	1
2023	Aguacateco	3
2023	Mam	349
2023	Q'anjob'al	44

Table 11

Post Policy Specific Mixtec Lable Use

Year	Language	Total Enrolled
2020	Mixteco, San Juan Mixtepec	3
2021	Mixtec, Ixtayutla	1
2021	Mixteco, San Juan Mixtepec	7
2022	Mixtec, Alcozauca	1
2022	Mixtec, Diuxi-Tilantongo	2
2022	Mixtec, San Juan Teita	1
2022	Mixtec, Santa Lucía Monteverde	1
2022	Mixtec, Southern Puebla	1
2022	Mixteco, San Juan Mixtepec	6
2023	Mixtec, Alcozauca	1
2023	Mixtec, Cacaloxtepec	1
2023	Mixtec, Diuxi-Tilantongo	2
2023	Mixtec, Northwest Oaxaca	1
2023	Mixtec, San Juan Teita	1
2023	Mixtec, Santa Lucía Monteverde	3
2023	Mixteco, San Juan Mixtepec	7

Table 12
Correlation Matrix

Variable	1	2	3	4
1. Spanish Enrollment	—	-.77	.69	-.51
2. Spanish ELL		—	-.51	.31
3. Other Count			—	-.24
4. Grouped Period				—

Note. None of the values are statistically significant

Table 13

Regression Discontinuity using proportion of total Spanish Enrollment

Predictor	Estimate	Std. Error	t-value	p-value
Intercept ($\beta 0$)	-0.0083	0.0082	-1.01	0.33
GroupedPeriod ($\beta 1$)	-0.00015	0.00013	-1.20	0.25
SpanishEnrollment ($\beta 2$)	0.071	0.0551	1.29	0.22

Figure 1
(Campbell, 2017)

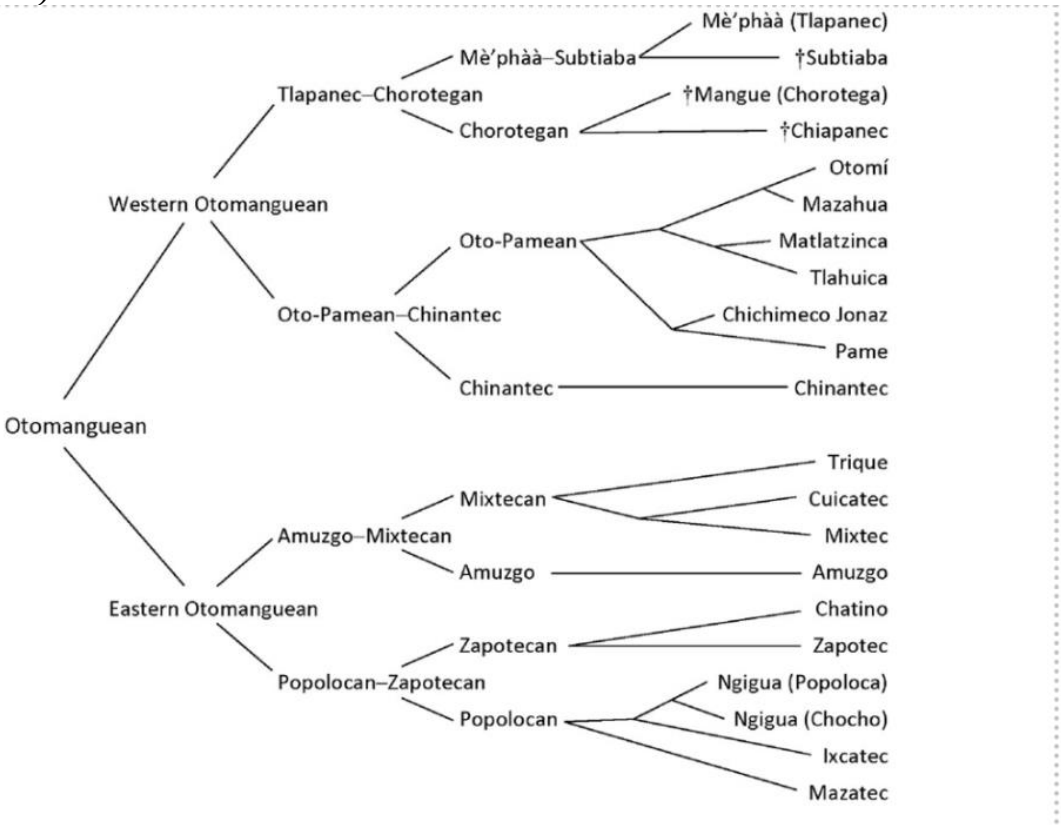


Figure 2
Mixtec branch from Otomanguean family (Hammarström et al., 2024)

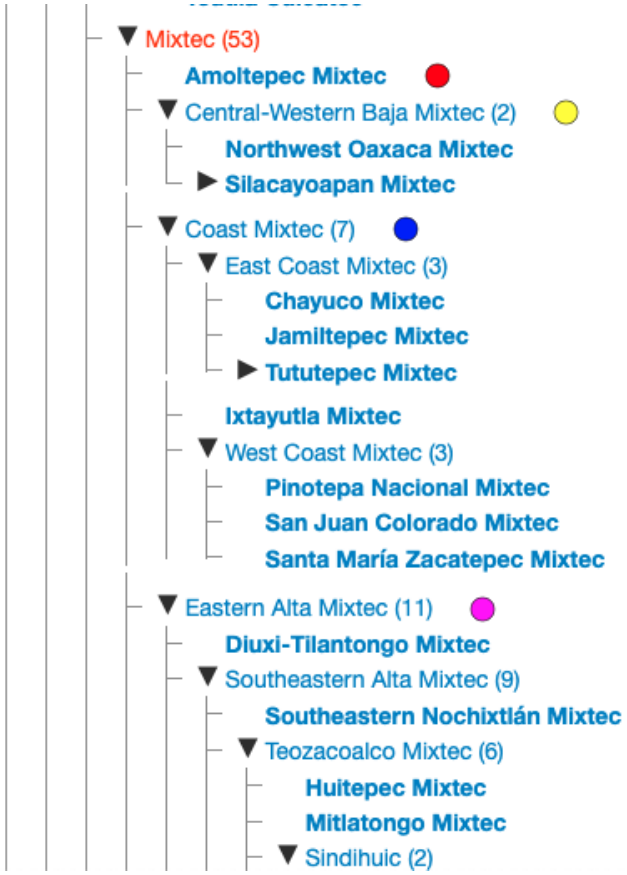


Figure 3

Zapotec Classification Details (Hammarström et al., 2024)

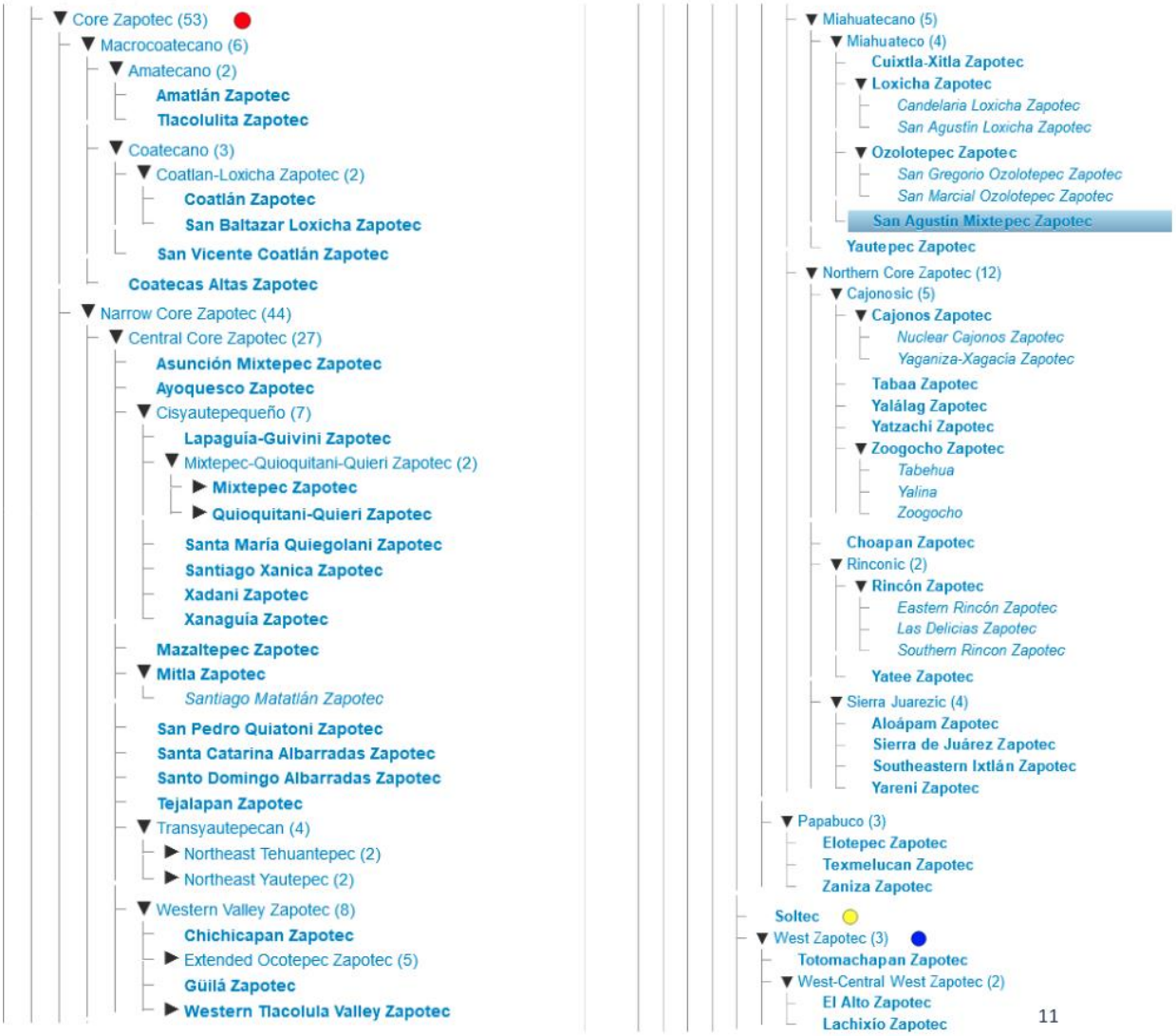


Figure 4
Top 10 Languages Enrolled in Oregon 2016-2023

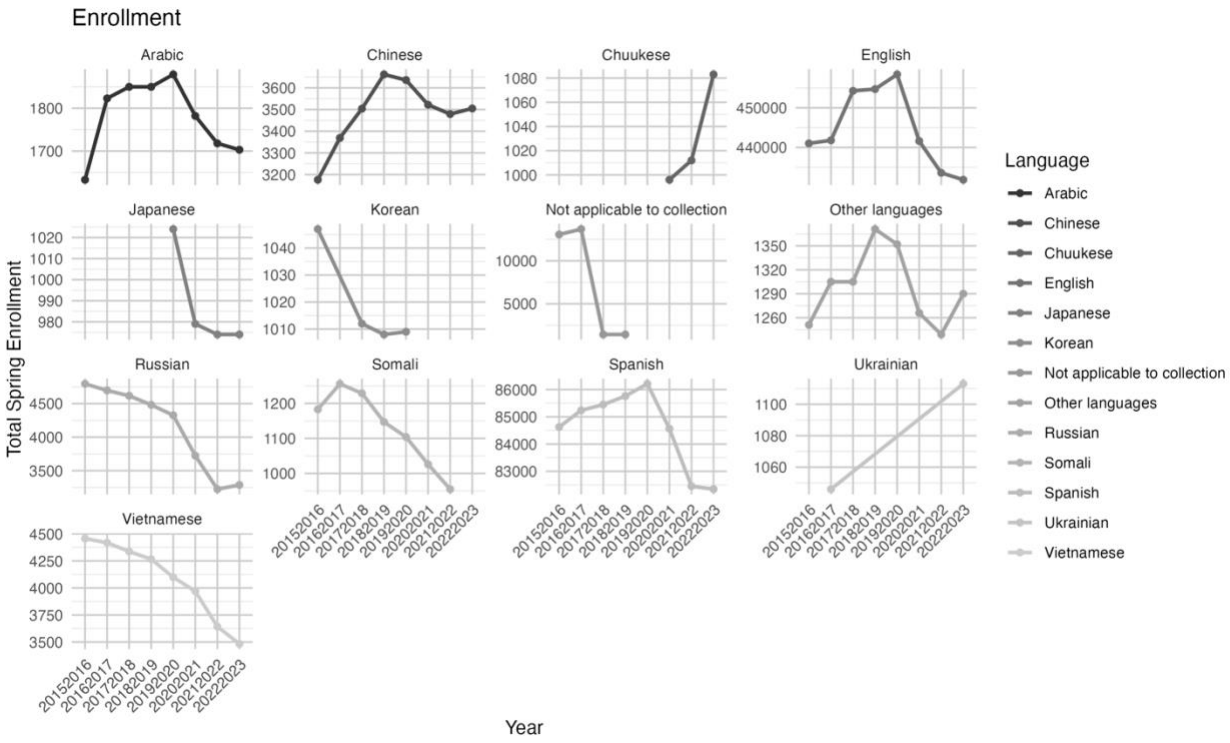


Figure 5

State Map Categorized by Language Guidance Coverage

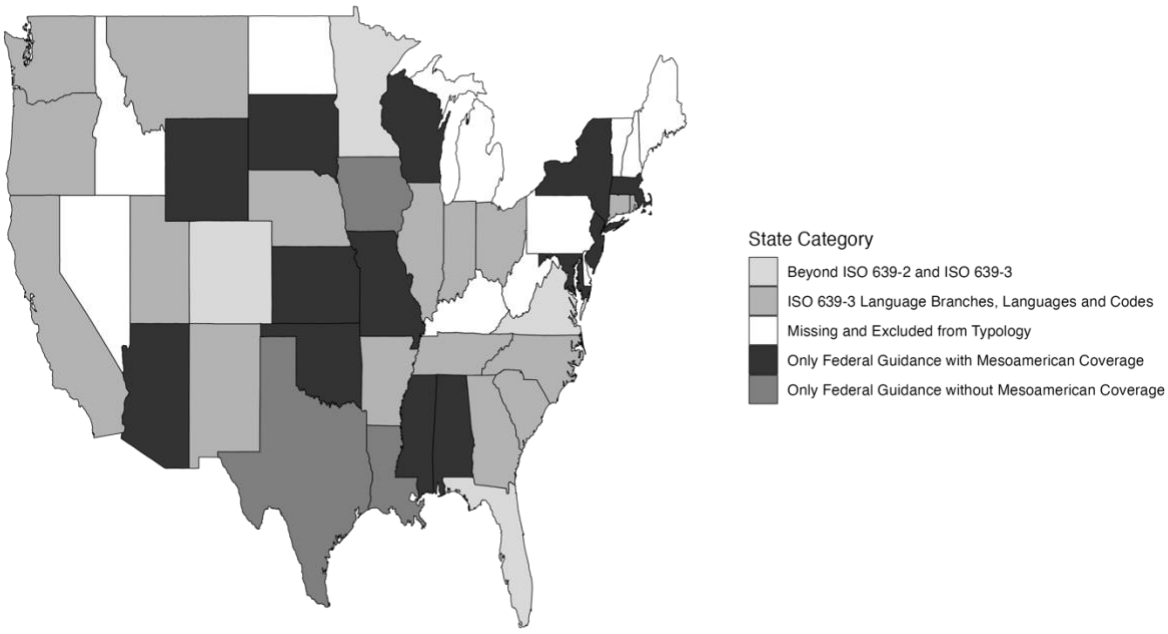


Figure 6

Languages Most Frequently Appearing across 43 States

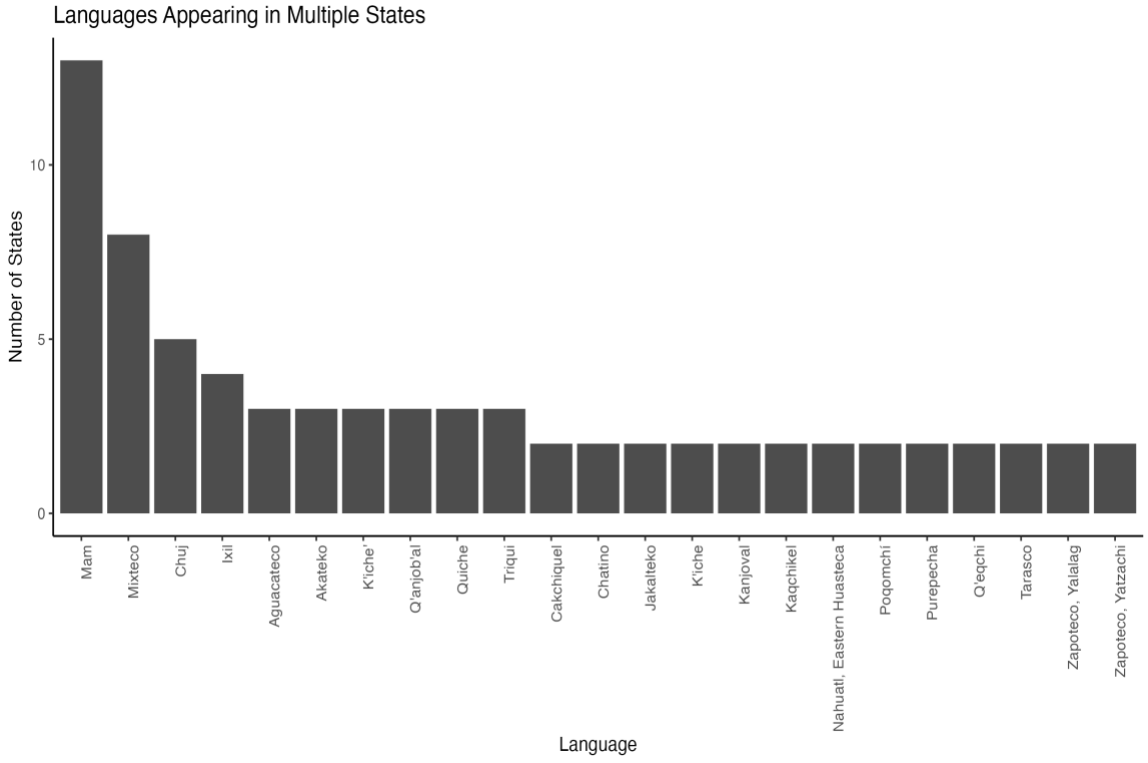
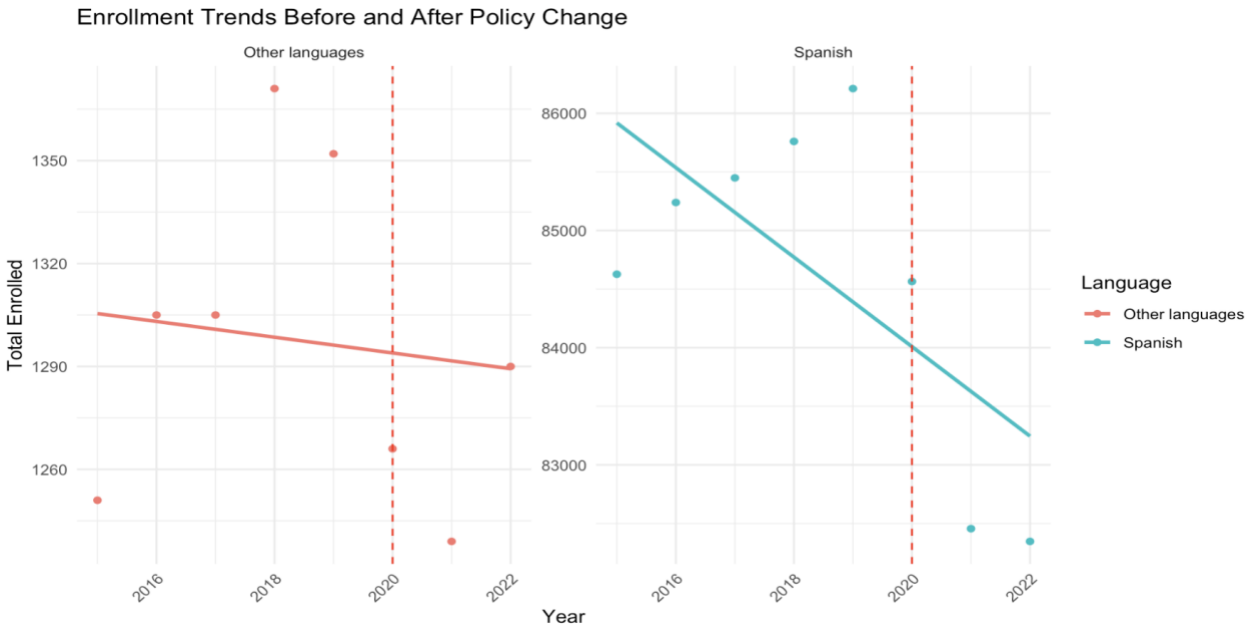


Figure 7
Regression Discontinuity Graph



Appendix A
Data Collection Sheet

State Name	Search Terms	URL	Date Accessed	Year of Document	Document Type	General Notes	More Notes/Impressions
Arizona	Arizona English Learners	Arizona OELAS Website	07/10/2024	2024	Other	This is the main website for EL-related things in Arizona	
Arizona	EL Data Tab	Arizona EL Data Tab	07/10/2024	2018	Language Code	Did not have an updated list. Will have to confirm that this is the most recent.	
California	Language Codes in CA SEA Search	CA SEA Language Codes	07/11/2024	2023-2024	Language Code	CDE receives requests from local educational agencies (LEAs) to add more languages to the CALPADS Language Code Set to better reflect the changing demographics	They explicitly mention both requests made by LEAs to CDE to add more languages, ISO 639-2, and ISO 639-3.

Appendix B
Example State Language Code Data Sheet

State	Reporting Codes Year	State Code Link	Coding Scheme	Zapotec Found?	Otomi Found?	Nahuatl Found?	Mixtec Found?	Mayan Languages Found?	Federal Code Adherence	General Impressions	Typology
California	30-May-24	CA Codes	Uses 3-letter combos	Yes	Yes	Yes	Yes, listed under "Otomian"	No individual languages	Full Adherence	Mixtec is miscoded under "Otomian." Few Mesoamerican languages listed.	Passive

Appendix C
Updated Coding Data Sheet Example

State	Adding Codes Beyond ISO 639-2	Uses Labels from ISO 639-3	Beyond ISO 639-3	Uses 9999 or unknown/unlisted
Mississippi	0	0	0	0
Colorado	1	1	1	0

Appendix D

Example Survey Questions

My name is Karen Zyskind, and I am a doctoral candidate at the University of Oregon's College of Education. I am currently conducting research on the language codes and language lists used to document the non-English languages spoken by English learners in the United States. My research focuses specifically on languages from Mesoamerica, a cultural and linguistic region that spans the southern half of Mexico and the northern third of Central America including all of Guatemala. There is a widespread assumption that people from Mesoamerica are Spanish speakers. However, there are several hundred Indigenous languages in Mesoamerican whose speakers are often invizibilized as they are labelled Hispanic or Latino. Of critical importance is how federal requirements influence or interact with state practices in documenting students' home languages. Your state has been identified as having language codes that recognize five or more Mesoamerican Indigenous languages. I would greatly appreciate your insights into the factors that influenced their recognition. If you are able to participate, your responses will help inform my understanding of how these language code additions were made.

Sample questions:

Question Block 1- General background questions

1. What is your current role/position within the State Department Agencey?
    a. Director
    b. Policy analyst
    c. Data
    d. Program Specialist
    e. Other (blank to be filled)

2. What state do you represent: Drop Down List to select
3. How familiar are you with your state's language codes and language lists for documenting non-English languages spoken by English Learners (ELs)?
    a. Very familiar
    b. Somewhat familiar
    c. Slightly familiar
    d. Not familiar at all

Question Block 2- Questions about Mesoamerican languages

1. Are you familiar with what Mesoamerican Indigenous languages are?
    a. Yes
    b. No

[If "No" is selected, display the following definition:]

Definition:

Mesoamerican Indigenous languages refer to the native languages spoken by indigenous communities in the Mesoamerican region, which includes parts of present-day Mexico, Guatemala, Belize, Honduras, and El Salvador. Examples of these languages include Mayan languages such as K'iche' and Yucatec), as well as Nahuatl, Zapotec, Mixtec, and Otomi languages, among many others.

2. How long has your state included Mesoamerican Indigenous languages in its language codes for English Learners?
    a. 1-3 years
    b. 4-6 years
    c. Not sure

d. Fill in text box

Question Block 3: Language codes

3. Does your state have specific policies or guidelines for adding languages to the language code lists used in your state student information system?

a. Yes

b. No

c. Other Fill in text box

Block 4: Please answer each statement based on your experience and knowledge. You may choose to elaborate on your responses in the optional comment box provided after each question.

9. Statement: “The federal policies and guidelines for adding languages to the language codes are clear and easy to follow."

a. Strongly Agree

b. Agree

c. Neutral

d. Disagree

e. Strongly Disagree

f. Comment Box [ ]