Journal of Educational Psychology

Language and Literacy Trajectories for Dual Language Learners (DLLs) With Different Home Languages: Linguistic Distance and Implications for Practice

Joshua Sussman, Karen Draney, and Mark Wilson

Online First Publication, May 18, 2023. https://dx.doi.org/10.1037/edu0000791

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

https://doi.org/10.1037/edu0000791

Language and Literacy Trajectories for Dual Language Learners (DLLs) With Different Home Languages: Linguistic Distance and Implications for Practice

Joshua Sussman, Karen Draney, and Mark Wilson
Berkeley Evaluation and Assessment Research (BEAR) Center, Berkeley School of Education, University of California, Berkeley

Using a large sample of longitudinal assessment data from children in publicly funded infant/toddler care, preschool, and kindergarten (analytic $N = 453{,}468$), this study modeled language and literacy trajectories from early infancy through kindergarten for dual language learners (DLLs) from homes representing many different languages and their peers from monolingual English-speaking homes. DLLs and their monolingual peers showed common developmental trajectories, including an initial period of rapid growth that slowed between 10 and 30 months, plateaued between 30 and 45 months, and then accelerated from 45 to 60 months of age. Although the general developmental patterns were similar, differences emerged between children in English-speaking versus non-English-speaking homes: DLLs' growth rates slowed more than their monolingual peers' between 10 and 30 months but then grew more sharply after 45 months. A linguistic distance hypothesis helped to explain the magnitude of the differences between groups of DLLs and their monolingual peers. The differences occurred as children passed through specific developmental levels of the Desired Results Developmental Profile (DRDP) assessment's learning progressions. The current study adds new quantitative information about language and literacy trajectories from early infancy through kindergarten among children from different language backgrounds and demonstrates how the findings may be applied to support effective instruction with culturally and linguistically diverse children.

Educational Impact and Implications Statement

This study examined developmental trajectories in language and literacy among children from different home language backgrounds in publicly funded infant/toddler care, preschool, and kindergarten programs. According to teachers' responses to developmental assessment rubrics, children showed similar trajectories of language and literacy development from infancy through 5 years, regardless of the language spoken at home. All children experienced an initial period of rapid growth that slowed between 10 and 30 months, held steady between 30 and 45 months, and then accelerated again from 45 to 60 months of age. Although the general patterns were similar, children classified as dual language learners grew less than their monolingual peers between 10 and 30 months but then grew more than their monolingual peers after 45 months. Specific developmental milestones were associated with the emerging and then waning differences, and greater linguistic similarity between a home language and English was associated with greater language and literacy development. Educators who apply their understanding of these trajectories, and the milestones at which they typically diverge, may have greater ability to identify children's difficulties, modify their instruction to target important developmental milestones, and better understand the degree to which children are progressing as expected.

Keywords: language development, dual language learner, developmental trajectory, learning progression, Desired Results Developmental Profile

Supplemental materials: https://doi.org/10.1037/edu0000791.supp

The United States is becoming an increasingly multilingual society. Recent decades have seen consistent increases in the population of children aged 0–8 who live in homes where a language other than English is regularly spoken. These children are often referred to as dual language learners (DLLs) in educational settings and the research literature. Recent statistics show that

Joshua Sussman https://orcid.org/0000-0002-8223-178X

This research was supported by a grant to Joshua Sussman from the American Educational Research Association (AERA) which receives funds for its AERA Grants Program from the National Science Foundation (NSF) under NSF award NSF-DRL (Grant 1749275). Opinions reflect those of the authors and do not necessarily reflect those of AERA or NSF.

Joshua Sussman served as lead for conceptualization, data curation, formal analysis, funding acquisition, methodology, visualization, writing–original draft, writing–review and editing. Karen Draney served in a supporting role for conceptualization, methodology, writing–review and editing. Mark Wilson served in a supporting role for conceptualization, methodology, writing–review and editing. The authors have no known conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Joshua Sussman, BEAR Center, Berkeley School of Education, University of California, Berkeley, 1995 University Avenue, Suite 120, Berkeley, CA 94704, United States. Email: jsussman@berkeley.edu

DLLs comprise one-third of all young children in the United States and 60% of those in California (Park et al., 2017). DLL populations are rapidly growing in the southern United States (Mancilla-Martinez et al., 2020) and are becoming more diverse nationwide: projections show immigration decreasing from Spanish-speaking countries and increasing from other countries, particularly those in Asia. Attention to effective education for the large and growing population of children classified as DLLs is clearly in the national interest (Fortuny et al., 2010; Heckman & Masterov, 2007).

The nation's early care and education (ECE) system is largely unprepared to respond to these demographic shifts and provide DLLs with access to high-quality, culturally, and linguistically responsive early learning environments (National Academies of Sciences, Engineering, and Medicine [NASEM], 2015; U.S. Department of Health and Human Services [DHHS] & U.S. Department of Education [DOE], 2016). Recent policy to increase ECE quality for all children has included the design and implementation of large-scale early learning and kindergarten entry assessment (KEA) systems (Ackerman, 2018a). This manuscript uses the broader term *early learning and KEA* because, in practice, KEAs are used in ECE and kindergarten settings for various purposes. One purpose is informing instructional decision making; a process often called formative assessment (e.g., Black & Wiliam, 2018; Sadler, 1989; Schildkamp et al., 2020). Barriers to effective formative assessment in ECE settings have been documented (Ackerman, 2019; Akers et al., 2016). Limitations in teachers' knowledge about (a) development in bilingual children and (b) the influences of culture and context on children's developmental trajectories are salient barriers that, if addressed, may support more effective formative assessment with DLLs (Dunphy, 2010; Gummer & Mandinach, 2015).

This secondary analysis of large-scale, longitudinal data from California's public infant/toddler, preschool, and kindergarten programs examines variation in children's language and literacy trajectories assessed using the current version of the Desired Results Developmental Profile (DRDP) assessment, the "DRDP (2015)," which is implemented statewide as the required assessment for all publicly funded ECE programs (though voluntary in kindergarten). In practice, children and teachers may use any language during DRDP assessment, though many children who are DLLs were assessed in English. The analysis used multilevel growth models to estimate children's developmental trajectories from early infancy through kindergarten and to study the variation across 23 different home language categories. We examined the role of linguistic distance for the five language categories with adequate data across the full age distribution. We interpreted the results in the context of a learning progression (Black et al., 2011; Heritage, 2008; Wilson, 2009): A set of developmental levels translated the statistical results into information about the specific developmental milestones that were most likely to be associated with changes in the growth trajectories. The discussion examines the potential value of the findings for supporting educators' understandings of language and literacy trajectories among children from different home language backgrounds, and how these understandings may support culturally and linguistically responsive practices.

Dual Language Learners (DLLs)

The term "dual language learners" is often used to describe children ages birth through 8 years who are learning two or more languages at the same time, or are learning a second language while continuing to develop their first (DHHS, 2008). However, a variety of labels are used in the literature and educational settings such as emergent bilingual, multilingual learner, limited English proficient, and others. Certain terms have been associated with deficit views of bilingualism (e.g., limited English proficient) whereas other terms reflect additive, strengths-based perspectives (e.g., emergent bilingual). In this report, we use the term DLL except when referring to the results of a study that uses a different term; in which case we use the term in the study. Researchers and educators also use a variety of procedures to classify children as DLLs, such as assessments of language proficiency, teacher reports, and parent/caregiver reports about the language(s) spoken at home. According to Surrain and Luk (2019), the various terms and procedures define heterogeneous populations and complicate efforts to synthesize conclusions across studies. Children in bilingual research samples are likely to differ along many important variables that influence learning and development, such as age of second-language acquisition, language balance, and various factors related to sociolinguistic context. Existing quantitative studies that avoid binary categorizations of DLL status and study linguistically diverse families (e.g., Winsler et al., 2014) show that disaggregating home language, as per the current paper, can produce nuanced and educationally relevant information about developmental processes and outcomes.

Early Language and Literacy Development in Children Who Are DLLs

The current consensus is that many factors influence language development in a child's native language, or L1 *and* their second language, or L2 (NASEM, 2017). Language development is cast as a social process driven by a child's exposure to language inputs and transactional experiences within multiple interacting contexts (Conboy, 2013). The *quantity* and *quality* of language input in L1 and L2 are important influences on language development (Paradis & Jia, 2017). This includes the total amount of language exposure (Hammer et al., 2014; Thordardottir, 2011) and various quality factors, such as parent/guardian language proficiency (in both L1 and L2), the purpose/context of child-directed speech, breadth of material that is read to the child, exposure to different types of media, opportunities to practice in different contexts, and quality of peer interactions (Escobar & Tamis-Lemonda 2015; Golinkoff et al., 2015; Hurtado et al., 2008; Schneidman et al., 2013). Broader sociocultural factors also contribute to early language development, such as immigration circumstances, family characteristics, and timing of L2 onset (Portes & Rumbaut, 2014). Emerging literature has also identified interactive, or cross-linguistic, influences on bilingual language acquisition (Hammer et al., 2020). The general idea is that a child's L1 development influences their L2 development and vice versa. An important conclusion from the literature is that variation in children's experiences with language drives variation in their language development.

Comparatively less research exists on young children's early literacy development, but the current consensus is that children's participation in literacy-rich environments and high-quality interactions is thought to promote their emergent literacy (see Hammer et al., 2014). A literacy-rich home environment offers children access and exposure to diverse print materials. Children's broader communities also play an important role as they can differ in the amount and diversity of available reading materials such as books and magazines (Neuman & Celano, 2001). Children can build emergent literacy through high-quality interactions with adults who read books to them in an emotionally positive manner and elicit their feedback during book reading, for example by use of dialogic ("wh") questions (NASEM, 2017). Research also points to cross-linguistic influences between emergent literacy in L1 and L2 (Tabors et al., 2001) and cross-domain influences between language and emergent literacy (e.g., Dickinson et al., 2010; Gardner-Neblett & Iruka, 2015; Rinaldi & Páez, 2008). The design of the DRDP assessment reflects these views; that children's early language and literacy development (in L1 and L2) are intertwined, and earlier skills are developmental precursors to later reading and writing abilities.

Linguistic Distance and DLLs' Language and Literacy Development

Linguistic distance is an emerging hypothesis for explaining, in part, the magnitude and ease of cross-language transfer (Floccia et al., 2018). The main idea is that the overlap between a child's L1 and L2 in aspects of oral language and writing may facilitate transfer and, ultimately, faster acquisition. Support for this hypothesis comes from bilingual education studies suggesting that instruction in Chinese, a relatively distant language from English, offers weaker support for English L2 development than instruction in Spanish, which has greater similarity with English (Valentino & Reardon, 2015; Wong-Fillmore et al., 1985). The literature contains studies that measure linguistic distance and test its association with important outcomes: Floccia et al. (2018) operationalized linguistic distance using a quantitative assessment of morphology, word structures, and grammar in linguistically diverse 2-year-olds. The authors found that smaller linguistic distance positively predicted language and early literacy outcomes in the child's "additional language," but not in English. Many studies dichotomize linguistic distance into binary groups for near distance (for similar languages) and far distance (for disparate languages): Blom et al. (2020) found that a near group of 24-month-old emergent bilinguals had similar Dutch receptive vocabularies as Dutch monolinguals, whereas the far group of bilingual children were behind the monolingual group (in Dutch). Using neuroscientific methods, Radman et al. (2021) found better reaction times in cognitive switching tasks in a near group of adults compared to a far group (language and nonlanguage switching tasks). In sum, the empirical studies consistently show associations between linguistic distance and relevant outcomes, but additional research is needed to resolve contradictory or incomplete findings and work toward more nuanced views of bilingualism. The current study tested the linguistic distance hypothesis with children who spoke five different home languages, using teacher report assessment data and a developmental lens spanning early infancy through kindergarten.

The Contributions of Quantitative, Longitudinal Research to Understanding DLLs' Language Development

Longitudinal research has contributed to evolving understandings about DLLs' language development. Historically, the literature has focused on the differences between Spanish-speaking DLLs and their monolingual English-speaking peers in outcomes such as expressive vocabulary, picture naming, word reading, and verbal memory (Choi et al., 2018; Hoff & Ribot, 2017; Jackson et al., 2014; Mancilla-Martinez & Lesaux, 2011; Páez et al., 2007; Vagh et al., 2009). A central conclusion from this research is that Spanish-speaking DLLs' English language skills lag behind those of their monolingual peers. However, the differences are not likely due to lower ability but rather differential opportunities to learn English and/or culturally biased conceptualizations of ability (e.g., measuring outcomes only in English). For example, Hoff et al. (2012) examined both English and Spanish vocabulary and grammar development in 47 high-SES Spanish–English DLLs and 56 non-DLLs aged 22–30 months. On average, DLLs' English development was behind that of the non-DLLs. However, DLLs' total vocabulary growth (L1 + L2) was equivalent to non-DLLs' growth in English. The conclusion was that DLLs develop as rapidly and as much as their non-DLL peers, but their opportunities to learn may be distributed over two languages.

The longitudinal research has also examined the sociolinguistic factors that can influence DLLs' language trajectories. Hoff et al. (2014) studied the association between parental L1 and expressive vocabulary (English and Spanish) in 31 children aged 22–48 months, finding that the group with two native Spanish-speaking parents showed larger gains in total vocabulary (L1 + L2) than the group with one native Spanish-speaking parent and one monolingual English-speaking parent. Ribot et al. (2018) examined the relationships between a child's language preference for switching from a native language to English and expressive vocabulary growth, finding that, after controlling for English input at home, their choice to switch to English was associated with a larger English vocabulary. These studies show that DLLs' opportunities to learn can depend on interactions between language context and individual factors.

A small number of longitudinal studies have examined language development in DLLs with home languages other than Spanish. MacLeod et al. (2019) studied French or English vocabulary growth in a 3.5- to 8-year-old sample of multilingual children with several home languages ($n = 106$) and their monolingual peers ($n = 211$). After controlling for child- and family-level variables, English/French bilinguals (labeled Language Majority children) had (a) similar rates of French or English vocabulary growth as the monolingual children and (b) greater rates of French or English vocabulary growth prior to school entry than the multilingual children who learned Italian, Arabic, Greek, Spanish, German, Portuguese, Polish, or Tagalog at home (i.e., Language Minority children). The distinction between minority and majority status explained the differences in French or English growth prior to school entry. A limitation is that children's growth in other home languages was not measured. Floccia et al.'s (2018) longitudinal study examined the association between linguistic distance and vocabulary in 2-year-olds learning British English and one of 13 other languages. After controlling for potential confounds (e.g., SES and language input), each of three measures of linguistic distance predicted vocabulary production and comprehension in children's non-English language, "showing, for the first time, that linguistic distance between the child's two languages predicts vocabulary outcome" (p. 1). In sum, as most longitudinal research with DLLs has focused on Spanish–English bilingual children, additional research with DLLs from diverse language backgrounds can help the field better understand variation in children's language development, including the potential influence of linguistic distance.

Translating Research to Practice: Effective Early Education for Culturally and Linguistically Diverse Populations

A strong research consensus states that, although bilingualism itself is a characteristic associated with neurological, cognitive, and cultural strengths and benefits (Kaiser et al., 2015; Murphey et al., 2014; Yow & Markman, 2011), the DLL population faces pervasive opportunity gaps in early childhood. This includes ECE teachers with inadequate preparation (Castro, 2014), education systems with limited cultural capacity (Montoya-Ávila et al., 2018), and culturally and linguistically biased assessment practices (Espinosa & Lopez, 2007; Ntuli et al., 2014). These problems may be exacerbated for DLLs with disabilities. It is widely agreed that supporting DLLs' strong English proficiency is a critical learning goal, but the nation's *de facto* system of English-only instruction (in part due to the lack of recruitment and retention of bilingual ECE teachers), runs counter to decades of research on the benefits of dual language models of ECE (e.g., Collier & Thomas, 2017). In addition, DLLs are more likely to live in poverty and belong to less educated families than monolingual children. The widely identified and discussed gaps in learning and development between DLLs and their peers that begin early and persist throughout the school are thought to be the product of these differential opportunities to learn (Aguirre-Muñoz & Amabisca, 2010).

Research demonstrates that high-quality ECE programs can support early learning and development for all children and may provide even greater benefits for children classified as DLLs (Sussman et al., 2022; Yazejian et al., 2015; Yoshikawa et al., 2013). However, a clear and consistent finding in the literature is that children classified as DLLs do not have the same access to high-quality programs as their monolingual peers (NASEM, 2017). Furthermore, DLLs' opportunities to learn can vary considerably across programs designed to be high quality (Jacoby & Lesaux, 2017). Aspects of quality that are thought to support DLLs' growth include strength-based views of bilingualism, strong family engagement, and developmentally, culturally, and linguistically responsive curriculum, instruction, and assessment (Banerjee & Luckner, 2014; Castro et al., 2011; National Education for the Association of Young Children, 2020). Cultural and linguistic match between families and educators (Gaitan, 2012; Markowitz et al., 2020) can support strength-based perspectives and family engagement. Using assessment data to support high-quality, individualized learning experiences, often called formative assessment, is considered especially critical for DLLs because their broader range of developmental experiences leads to increased needs for differentiation in the classroom (Alvarez et al., 2014; Davison, 2019; Heritage et al., 2012). Indeed, the current report aligns with research and theory showing that early childhood educators can use developmentally informed assessments that point to a child's current skills, and their likely next steps, to design effective and inclusive learning environments (see Mangione et al., 2019; National Research Council [NRC], 2008). Next, we describe our assessment—the DRDP—that was designed for these purposes. But first, we briefly discuss federal policy that guided the recent construction of early childhood assessment systems.

Supporting High-Quality Early Care and Education (ECE) With Early Learning and Kindergarten Entry Assessments (KEAs)

Aiming to address the need for valid and reliable assessments in ECE settings and to support high-quality ECE for all children, federal initiatives such as the National Education Goals Panel (National Education Goals Panel [NEGP], 1995) prioritized efforts to define the critical areas of early learning and development that support school success. Later, the federal Race to the Top—Early Learning Challenge (RTT-ELC) grant program and Enhanced Assessment Grants supported state-managed efforts to design and implement dozens of large-scale assessments, commonly referred to as KEAs (Ackerman, 2018a; Weisenfeld et al., 2020), to assess these critical areas. The assessments are used for various purposes: for accountability (e.g., measuring and tracking the number of children considered "ready" for school), to inform instructional practices with individual children, for formative evaluations that guide program decisions, and as variables in research.

The RTT-ELC legislation stipulated that the assessments must be valid, reliable, and fair for children with disabilities and limited English proficient students. Many states designed and implemented observational, or "authentic," assessments, that rely on extended teacher observations of children instead of the "direct" assessments of student proficiency that generally rely on normative interpretations and have been implicated in deficit views of bilingualism (see Atkins-Burnett, 2007; Castro & Artiles, 2021 for discussions about issues with direct assessments). Ackerman (2018b) established criteria for evaluating the validity of KEAs with kindergarteners classified as English learners (ELs), who are commonly regarded as children whose English skills are not yet developed enough to permit their access to English language curricula. The criteria focused on two assessment purposes: accurately documenting children's knowledge and informing instruction. The validity criteria included the degree to which assessments contained EL-specific content at the item level (i.e., accurate and culturally relevant item-level translations and exemplars), linguistic accommodations (e.g., allowing students to use any language and nonverbal cues), and other features such as evidence of psychometric reliability and measurement invariance for ELs. Ackerman evaluated nine purposely selected KEAs and their materials, placing the assessments into three major groups within a less to more continuum based on the degree to which evidence supported the use of the measures with ELs. Among the nine instruments, only two assessments—both versions of DRDP (California and Illinois)—met the criteria for the highest category of linguistic support. The conclusion was that DRDP provides a robust set of EL-specific supports and has a portfolio of EL-relevant reliability and validity studies. NASEM (2017) recognized that, given the limited assessment options, the DRDP is a suitable tool for use with DLLs, although additional validation work is sorely needed.

The Desired Results Developmental Profile (DRDP) Assessment

The DRDP is an observational measure: Teachers use rubrics to rate children's observable behaviors based on early learning standards (called "foundations" in California). To inform their ratings, teachers collect evidence through observation and documentation. A teacher's ratings reflect assessment in NEGP's (1995) five essential domains as well as several optional domains for History & Social Science, Visual and Performing Arts, English Language Development, and Spanish Language Development. The DRDP is separated into three versions, called age levels, for use in (a) infant/toddler group care settings, (b) preschool classrooms, and

(c) kindergarten classrooms. The three age levels are psychometrically linked into a single measurement continuum (Karelitz et al., 2010). Brief DRDP versions, called Views, are also available (see Supplement 1 in the online supplemental materials or Draney et al., 2022, for more information). This study uses the complete version of the instrument, named the Comprehensive View. An online DRDP data system displays child-level reports that describe children's strengths.

The DRDP was designed to serve the needs of multiple users at different levels of education (see NRC, 2008; Stiggins, 2008). In practice, DRDP results are summarized at the child, classroom, program, district, and statewide levels for continuous improvement and planning for curriculum, instruction, and school environments (Yun et al., 2021). Evaluations use DRDP data to examine the impact of agency-wide efforts to strengthen children's learning and development (e.g., McCarthy et al., 2020). Evidence from districts suggests that the DRDP has had positive effects related to systems change: DRDP has been used to vertically align developmentally appropriate curriculum and instruction in preschool, kindergarten, and first grade; has been described as a centerpiece for family discussions about what children know and can do; and has been used to inform strategic district initiatives aimed at boosting child development in specific content areas (Yun et al., 2021). Several small-scale qualitative studies based on reports from teachers and program administrators indicate that DRDP helps teachers better understand their students (Illinois State Board of Education, 2020), benefits teaching and instructional planning (Cannady et al., 2019), and improves program quality and teachers' knowledge and instructional skills (Khalegi, 2010).

Relating Assessment Scores Back to the Early Learning Standards: A Learning Progressions Approach

The DRDP was designed using learning progressions (Duschl et al., 2011; Heritage, 2008; Wilson & Lehrer, 2021) in the context of a construct modeling (Wilson, in press) approach to assessment design. The learning progressions are part of an assessment model that is used to translate DRDP's conventional scaled scores into information about children's observable developmental skills, capacities, and behaviors. Learning progressions are commonly described as successively more sophisticated ways of thinking and behaving that tend to follow one another as children mature and learn. As is typical, the DRDP's item-level learning progressions were designed to be granular interpretations of learning standards: they are grounded in developmental research, organized around central ideas and milestones in child development, and aim to operationalize what children need to know and be able to do to progress through the standards as their ideas and skills develop over time (Mangione et al., 2019).

The DRDP's learning progressions for language and literacy development encompass 10 areas, including language comprehension, expression, phonological awareness, and interest in literacy (see Supplement 1 in the online supplemental materials for additional details). Each area is a DRDP item. The three DRDP age levels contain different sets of developmentally appropriate items. For example, all age levels contain language items, whereas assessment of most literacy skills begins with the Preschool age level. The additional literacy items for older children assess understanding of print concepts, reading comprehension, letter and word knowledge, and emergent writing.

Each DRDP item has its own observational guide that operationalizes the steps in learning and development within a domain. Figure 1 contains the Preschool age level observational guide for the Communication and Use of Language (Expressive) item. First, the major steps in the learning progression are summarized as four major categories of development (Responding, Exploring, Building, and Integrating) and nine successive developmental levels intended to span the range of preschool milestones. Second, the developmental levels contain summary definitions and examples that describe what it means to progress toward more mature understandings and performances. The DRDP manual (https://www.desiredresults.us/drdp-forms) contains rich descriptions and exemplars for each item and developmental level.

Importantly, learning progressions are not considered stage models, fixed, or immutable. They articulate what are thought to be regular and predictable sequences of development in most children, such as those aspects of language development that tend to progress similarly for bilingual and monolingual children (see Hammer et al., 2014). Learning progressions should be seen as general tendencies rather than strict orders. Learning emerges in response to a curriculum, slippage is typical, and different cultural and linguistic environments can support and ascribe different meanings to behaviors (e.g., Saxe, 1988).

Pedagogically Relevant Assessment Scores

A child's location within a item-level learning progression is an indicator of their current skills and capacities with implications for their likely next step(s). However, with over 50 items in the full version of the DRDP, it is usually impractical to examine each child's location on each individual item—especially for those who seek to understand broad trends across children (e.g., decision makers at the program and policy levels). Aiming for an efficient means of providing pedagogically relevant information, the DRDP's psychometric framework translates a child's scaled score in a domain, such as Language and Literacy Development, into a developmental level for that domain. The assessed developmental level implies a child's position wihtin the item-level learning progressions. Figure 2 contains a schematic of this procedure and its products: The horizontal line in Part A represents the child's DRDP score along a unidimensional continuous latent variable, created by scaling raw scores from items within a domain using a polytomous Rasch model (Masters, 1982). Part B shows that the child's scaled score of −4.8 logits falls within the Exploring Earlier (EE) developmental level. Part C shows that a child assessed at the EE developmental level is expected to be at the EE level for most or all items within the domain (though this is not always the case—see below). Knowing that a child is at the EE level, a teacher may efficiently scan the developmental milestones across all items within the domain to understand and summarize the child's development. Similarly, a teacher or administrator may examine the distribution of developmental levels in a group of children. To create the developmental levels in Part B, a Rasch-based standard-setting method in the item mapping family (Diakow et al., 2013; see also Draney & Wilson, 2011) was used to partition the latent variable's distribution into 11 successive levels. The width of each level corresponds to the relative amount of development that is likely to result in a child passing through the level. The widths were set by the item parameters that account for knowledge and skills that emerge earlier or later in development

Figure 1

An Excerpt From the Desired Results Developmental Profile Assessment

Developmental Domain: LLD — Language and Literacy Development

LLD 3: Communication and Use of Language (Expressive)

Child's communication develops from nonverbal communication to using language with increasingly complex words and sentences

Mark the latest developmental level the child has mastered:

Responding		Exploring			Building			Integrating
Earlier	Later	Earlier	Middle	Later	Earlier	Middle	Later	Earlier
Makes sounds spontaneously	**Uses sounds, gestures, or facial expressions to communicate**	**Uses a few "first words," word-like sounds, or gestures to communicate**	**Uses a variety of single words to communicate**	**Uses two words together to communicate**	**Uses short phrases or sentences of more than two words to communicate**	**Uses short sentences that contain nouns, verbs, and other words, such as adjectives and recently encountered vocabulary, to communicate**	**Uses phrases and sentences with a variety of word forms, including past tense, future tense, plurals, pronouns, or possessives, to communicate, sometimes with errors**	**Combines phrases and sentences with a variety of word forms to communicate ideas or to describe people, objects, or events**
Possible Examples								
• Cries. • Coos. • Gurgles.	• Smiles when a familiar person approaches. • Cries or looks at an adult when hungry. • Vocalizes or babbles while interacting with an adult.	• Asks for food when hungry, by using a special word, sound, or gesture for food. • Communicates, "Mama," "Dada," "Baba," or similar word approximations. • Reaches for or gestures for an object.	• Names familiar foods, toys, or family members. • Communicates ideas such as "No," "More," or "Up." • Indicates a picture of a ball when asked what the child wants to play with next.	• Communicates, "Mommy come," when wanting a parent. • Communicates, "More juice," when thirsty. • Communicates, "我的卡車!" ["My truck!" in Chinese] after another child takes a toy truck.	• Communicates, "A mí me toca," ["It's my turn," in Spanish] when an adult brings the pet rabbit for a visit. • Communicates, "I want mommy." • Communicates, "I like dogs," while looking at an animal book.	• Communicates, "The rabbit is scared," when the pet rabbit snuggles into an adult's lap. ("Scared" is an adjective.) • Communicates using a communication board, "I need a tissue. My nose is runny." ("Tissue" is a noun and "runny" is an adjective.) • Communicates, "Malaking malaki ang aso namin," ["Our dog is huge," in Tagalog] after hearing a peer use the word "huge." ("Huge" is a recently encountered vocabulary word.)	• Communicates to a peer, during play, "Yesterday we made vegetable soup." ("We" is a pronoun; "made" is past tense.) • Communicates, "His birthday is tomorrow. He will be five." ("His" is a possessive pronoun; "he" is a pronoun; "will be" is future tense.) • Communicates in sign language that the cat's feet are wet. ("Cat's" is possessive; "feet" is plural.) • Communicates, "He runned really fast," [He *ran* really fast]. ("Runned" is past tense with a grammatical error.)	• Communicates, "Dragons don't need bikes 'cause they can fly. They have really big wings." • Communicates to a peer, "Let's hurry and clean up so we can go outside to ride bikes." • Communicates via spoken words, signs, or a communication device, "The dog ate the cat's food, and then he got in trouble. We put him outside and he was very sad." • Communicates, "Mi abuela es muy vieja. Tiene el cabello blanco y muchas arrugas," ["My grandma is really old. She has white hair and lots of wrinkles," in Spanish].

Note. From *DRDP (2015): An Early Childhood Developmental Continuum: Preschool Comprehensive View* (p. 15), by the California Department of Education, Early Learning and Care Division and Special Education Division, 2015. Copyright by the California Department of Education. Reprinted with permission. See the online article for the color version of this figure.

and the time in-between the emergence of developmental skills and behaviors. Sussman et al. (2021) used a classification accuracy approach to examine the agreement between a child's developmental level and each of their item levels for the five major DRDP domains. In a sample of 1.5 million assessments (with some children observed more than once), the authors found an average of 69% agreement between the developmental (domain) level and the item level. When the criterion was relaxed to include adjacent categories, agreement climbed to 94%. Although the developmental level is not a substitute for teacher judgment, it is a psychometrically accurate and efficient source of information to support understanding, communication, and planning. The current study is the first published account of the use of DRDP's developmental levels as a lens to analyze children's developmental trajectories.

Current Study

This longitudinal analysis explores variation in young children's language and early literacy trajectories for three main purposes: (a) to test the degree to which a linguistic distance hypothesis explains variation between home languages; (b) to explore how the results may be applied to support more effective education for children from culturally and linguistically diverse families; and (c) to evaluate the cross-linguistic applicability of the DRDP's developmental levels and learning progressions. This secondary analysis of large-scale, longitudinal assessment data from public ECE and kindergarten programs in California used multilevel growth curve analyses to examine children's language and literacy trajectories. We focused the analysis on how children's growth trajectories progressed through the DRDP's developmental levels and to what extent home language moderated growth. Initially, a model-building process established a suitable model of development from early infancy through kindergarten. Then, the associations between children's developmental trajectories and 23 home language categories were explored. The developmental trajectories were interpreted in the context of a learning progression that translated quantitative results into information about children's developmental milestones. A linguistic distance hypothesis was tested for the five home languages with adequate data across the entire early infancy through kindergarten age distribution: Arabic, Cantonese, English (monolingual), Mandarin (Putonghua), and Spanish. According to the linguistic distance hypothesis, the estimates for children who spoke Spanish were predicted to be greater than those who spoke Arabic or a Chinese language. Compared to the existing quantitative literature, the current study aims to describe children's development and variation over a very wide age span (early infancy through kindergarten), with precision (via a large, diverse sample and polynomial regression), and with pedagogical relevance (via interpreting the results using the DRDP's developmental levels).

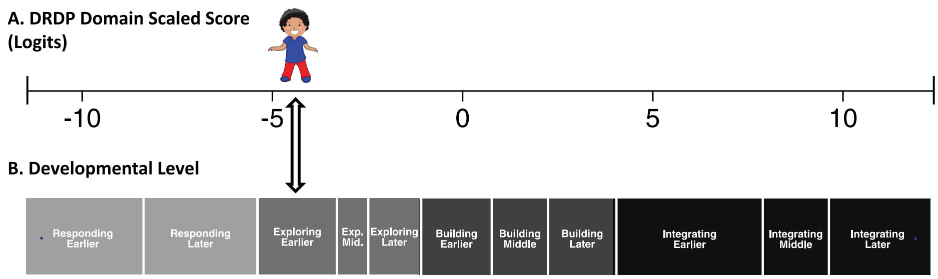
Method

Participants

The full sample contained 1,035,370 observations of 519,142 children within 4,543 programs who were rated by their teachers using the DRDP (Comprehensive) between Fall 2015 and Spring 2018. The final regression model used an analytic subsample of 843,349 observations from 453,468 children. Importantly, the

Figure 2

Connections Between Desired Results Developmental Profile (DRDP) Scaled Scores, Developmental Levels, and Learning Progressions



C. Learning Progression (Item) Level

Language and Literacy Development Item (Infant/Toddler Age Level)	Developmental Milestones in the Exploring Earlier Level
LLD1: Understanding of Language (Receptive)	Recognizes a few frequently used words or gestures in familiar situations
LLD2: Responsiveness to Language	Responds to a few frequently used words or gestures in familiar situations
LLD3: Communication/Use of Language (Expressive)	Uses a few "first words," word-like sounds, or gestures to communicate
LLD4: Reciprocal Communication and Conversation	Engages in brief back-and-forth communication with a familiar adult, using word approximations, vocalizations, gestures, or facial expressions
LLD5: Interest in Literacy	Attends briefly to a familiar adult reading books, singing songs, or saying rhymes

Note. Adapted from *DRDP (2015): A Developmental Continuum From Early Infancy to Kindergarten Entry: Infant/Toddler View for Use With Infants and Toddlers,* by the California Department of Education, Early Learning and Care Division and Special Education Division, 2015, and from *DRDP Technical Report for Early Infancy Through Kindergarten: Structural Validity and Reliability,* by K. Draney, J. Sussman, P. Gochyyev, K. Kriener-Althen, E. K. Newton, & P. Mangione, 2022. Copyright by the California Department of Education. Adapted with permission. See the online article for the color version of this figure.

participants attended publicly funded infant/toddler, preschool, and kindergarten programs in California and are representative of a culturally and linguistically diverse population of children from low-income families (not the U.S. population). Most children met California's Title 5 income-eligibility requirements, with priority given to children who were considered at risk and those with greatest financial needs. Data on children's socioeconomic status were not available to analyze, but the sample represents a low-income SES stratum. The DRDP is state mandated in publicly funded infant/toddler and preschool programs, so the sample approaches measurement of the entire public ECE population in California.

Children's demographics were reported by their teachers (or support staff) using online software (see Supplement 1 in the online supplemental materials for more information about demographic data collection). Demographic information refers to the full sample unless otherwise noted. The children were 3–80 months of age at assessment. The grand mean age across observations was 50.1 months ($SD = 11.5$). Children were assessed between 1 and 9 occasions ($M = 2.0$, $SD = 1.03$). Thirty-four percent of children were assessed once, 45% twice, 13% thrice, and 8% four or more times. Ninety-eight percent of the sample attended a single program, hence the effect of child mobility at Level 3 (i.e., switching programs) was ignored. Answers to a set of questions determined each child's DLL status, home language, whether a teacher spoke the child's home language with the child (i.e., language match), and other demographic information. Children were classified as DLLs based on the teacher's yes or no answer to the following question: *Is a language other than English spoken in the home?* Teachers were also asked to list the child's home language(s) using seven options (*English*, *Spanish*, *Vietnamese*, *Cantonese*, *Hmong*, *Tagalog/Pilipino*, and *Other [specify]*). Children with more than one non-English language were classified as Multilingual. A language match between teacher and child was determined to occur if the teacher reported using the child's home language with the child (teachers were asked to report *Language(s) used with this child*). Teachers were also asked to report the type of childcare setting: Table 1 shows the levels of the setting variable. State school specifically refers to the California State Preschool Program (CSPP). Finally, teachers were asked whether a child was in special education (had an Individualized Education Plan or Family Support Plan) and the child's race/ethnicity and sex.

Table 1 contains the distribution of teacher-reported variables in the sample, cross-tabulated with DLL status. Overall, 51% of the sample were classified as DLLs. The top section of the table contains variables that do not change within children (race/ethnicity and sex), and *n* reflects the number of *participants*. As expected, racially/ethnically diverse students were overrepresented among DLLs. The bottom section of the table contains variables that may change both within and between children over subsequent measurement occasions (i.e., childcare setting, teacher/student language match, and special education status) and lists the total number of *observations*. Thirteen percent of children had a childcare setting that changed. A child's setting could change, even if they remained in

Table 1
Sample Size by Demographic and DLL Status

	DLL		Non-DLL	
Demographic variable	*n*	%	*n*	%
Observations that vary only between children (*n* = children)				
Race/ethnicity				
Asian/Pacific Islander	23,929	9.0	13,389	5.3
Black	3,180	1.2	43,641	17.2
Hispanic	218,246	82.0	124,835	49.3
White	9,984	3.8	53,170	21.0
Other	9,220	3.5	17,157	6.8
Missing	1,493	0.6	898	0.4
Total	266,052	100.0	253,090	100.0
Sex				
Female	133,125	50.0	125,760	49.7
Male	132,798	49.9	127,234	50.3
Missing	129	<0.1	96	<0.1
Observations that may vary within and between children (*n* = observations)				
Childcare setting				
Center-based	35,119	6.5	41,533	8.4
Family care	10,508	2.0	11,083	2.2
Head Start	139,627	26.0	128,132	25.8
State school	274,513	51.0	242,208	48.7
Other	16,770	3.1	18,910	3.8
Missing	61,487	11.4	55,480	11.2
Total	538,024	100.0	497,346	100.0
Language match				
Yes	242,103	45.0	497,346	100
No	292,504	54.4	0	0
Missing	3,417	0.6	0	0
Special education				
Yes	29,609	5.5	27,527	5.5
No	457,945	85.1	426,377	85.7
Missing	50,470	9.4	43,442	8.7

Note. DLL = dual language learner.

the same program, because of program-level changes (e.g., a child care center was awarded a state school contract) or child-level changes (e.g., a change in family eligibility for state school resulted in the child switching to a different class in the same program). Seven percent of children had a language match that changed, and 5% had a special education status that changed. The number of DLLs and non-DLLs was balanced across sex, setting, and special education status. The preponderance of children in State School and Head Start settings reflects the nature of public preschool opportunities in the state. Forty-five percent of DLLs' teachers reported speaking the child's home language (i.e., a language match).

Table 2 contains the distribution of home languages in the sample (including English) where *n* equals the number of children. The original language variable contained 64 languages, but those with fewer than 200 children were reassigned to the "Other" category. Many languages were represented in sufficient number for growth modeling, but as shown in the next section, there were relatively few assessments completed for very young children. A child's home language may not vary over time: discrepancies across different measurement occasions (which occurred in 8% of children) were resolved by setting the child's home language to their final observation, when a teacher would be expected to know the child best. However, results allowing language to vary within children were essentially identical.

Table 2
Sample Size by Home Language

Home language	*n*	%
Arabic	3,570	0.69
Armenian	933	0.18
Cantonese	4,450	0.86
Chaldean	679	0.13
English (non-DLL)	253,090	48.75
Farsi	1,219	0.23
French	236	0.05
Hindi	570	0.11
Hmong	1,341	0.26
Japanese	263	0.05
Khmer	499	0.1
Korean	1,072	0.21
Mandarin (Putonghua)	1,642	0.32
Multilingual	533	0.1
Pashto	213	0.04
Portuguese	397	0.08
Punjabi	2,072	0.4
Russian	574	0.11
Spanish	209,549	40.36
Tagalog	1,239	0.24
Urdu	369	0.07
Vietnamese	4,473	0.86
Other	27,146	5.23
Missing	3,013	0.58
Total	519,142	100

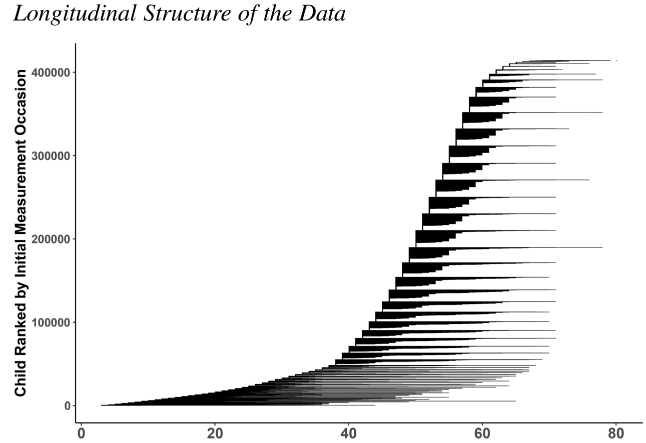
Note. DLL = dual language learner.

Data

Longitudinal Structure of the Data

The longitudinal dataset has varying assessment occasions (assessments at varying points in time) and unequal spacing between occasions (time intervals between adjacent occasions were not constant over time or across children). Figure 3 depicts the data's longitudinal structure in the analytic sample for the final model. The graph contains

Figure 3
Longitudinal Structure of the Data



Note. A rough visualization of the longitudinal structure of the data containing 453,468 horizontal lines, one line per child, where the beginning of each line shows the child's age at their first assessment and the end of each line shows the child's age at their final assessment. Apparent vertical lines denote children with single assessments.

453,468 horizontal lines; one per each child in the analytic subsample. Each line tracks the assessment window (first to final) for a child as a function of their age. The child IDs on the *y*-axis were sorted by initial age to show stairsteps of cohorts—children who received their first assessment at the same age. Each line's horizontal length tracks the number of months that elapsed between the first and the final assessment. Some details are obscured: the number of assessments between the first and last cannot be discerned. The apparently vertical lines at the front of the stairsteps denote children with single assessments, but such children cannot be seen prior to about 40 months (the steps are more densely packed). Despite these limitations, the graph accurately shows that (a) the longitudinal data are an amalgam of overlapping assessments, and (b) the amount of information available for estimating growth trajectories changes with age. DLL status was balanced across different ages. Supplement 2 in the online supplemental materials contains information about the data structure for children with different home languages.

Individual Growth Trajectories

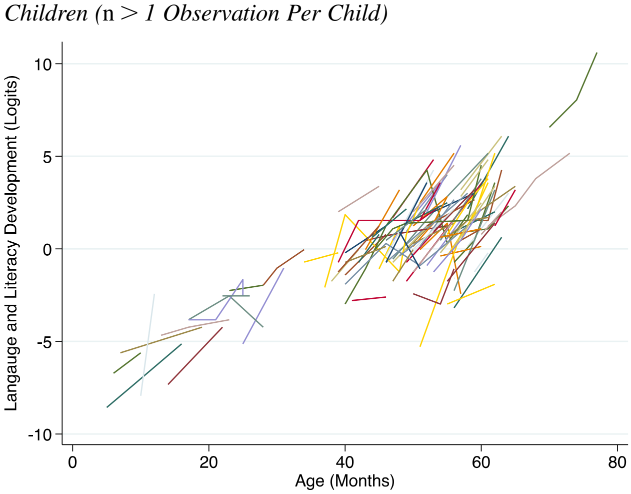
It is common to plot raw data to examine variation and justify the functional form of the modeled trajectories. Figure 4 contains a spaghetti plot with 150 randomly sampled Level 2 clusters (children), ignoring cases with only one observation. Each line graphs the changes in a child's assessment scores from their first through their final assessment. Each child was assessed between 2 and 4 occasions. The graph shows steeper slopes in the 40- to 60-month range than earlier months, suggesting an acceleration in learning and development around this time. The graph also shows the possibility of an increase in variance with age, which is allowed for within the model and does not threaten the accuracy or consistency of the estimates.

DRDP Age Levels

As mentioned in the introduction, the current DRDP is a set of three vertically scaled instruments designed to measure children's learning and development in three age levels: infants and toddlers, preschoolers, and kindergarteners. Table 3 contains descriptive

Figure 4

Spaghetti Plot of Trajectories for 150 Randomly Sampled Children (n > 1 Observation Per Child)



Note. See the online article for the color version of this figure.

Table 3

Distribution of Language and Literacy Scaled Scores (in Logits) for Different DRDP Views

Age level	*M*	*SD*	Min	Max	*n*
Infant/toddler	−3.23	2.40	−10.64	0.07	108,815
Preschool	1.03	2.34	−10.64	6.13	914,735
Kindergarten	3.97	3.02	−0.17	12.99	11,810
Missing	−0.05	2.09	−3.65	3.01	10
Total	0.62	2.72	−10.64	12.99	1,035,370

Note. DRDP = Desired Results Developmental Profile.

statistics for the DRDP outcome, separated by age level, for the 519,142 children in the full sample. Assessments without an age level were marked as missing. Importantly, developmental trajectories may be influenced by measurement artifacts of the different instrments (e.g., floor and ceiling effects). Hence, the trajectories reported in the results, based on the full data, were replicated using censored data samples: trajectories from 0 to 36 months were replicated using data from the infant/toddler age level only, and the trajectories from 0 to 60 months were replicated without the kindergarten age level. In all cases, the results were closely comparable between the full and censored datasets. Qualitatively, the trajectories were indistinguishable across datasets and quantitative differences had negligible practical significance.

DRDP Validity, Reliability, and Fairness

An assessment's validity is for a specific purpose, and the DRDP serves multiple purposes, including (a) measurement of individual children's development; (b) teacher reflection and instructional planning; (c) collaboration with families; (d) support for aligning curriculum, instruction, and assessment across school levels (infant/toddler, preschool, and elementary school); (e) a context for teacher professional development; and (f) data use at the program and policy levels. Decades of DRDP development, validation, and use have produced a corpus of information: descriptions of alignment with California's early learning standards (called Learning Foundations), standard-setting activities, examinations of cultural and linguistic responsiveness, and other reliability and validity evidence. An exhaustive review is beyond the scope of this paper, but the annotated bibliography in Supplement 3 in the online supplemental materials contains many details. In addition, Supplement 1 in the online supplemental materials contains a summary of the DRDP training program that includes online modules, a YouTube channel, webinars, group training, one-to-one consultations, and coaching. The remainder of this section summarizes salient reliability and validity evidence.

First, interrater reliability is crucial for an observational assessment. The DRDP has a portfolio of studies conducted across various DRDP generations that show adequate to strong interrater reliability (see Supplement 3 in the online supplemental materials). A recent study calculated Cohen's weighted κ values between 0.77 and 0.88 for pairs of infant/toddler raters, 0.68–0.81 for pairs of preschool raters, and 0.40–0.68 for pairs of kindergarten raters (WestEd, 2022a). The lower agreement between kindergarten raters did not match prior studies showing similar agreement across age levels and was attributed to the use of secondary raters who were unfamiliar with the children. Second, a potential criticism levied against the DRDP (Stipek, 2019) is that the correlations between children's scores in multiple dimensions are "too high," and this may indicate that the assessment lacks discriminant validity. However, WestEd (2022b) documented patterns of convergent evidence (moderate correlations with direct assessments designed to measure similar skills) and discriminant evidence (lower correlations with dissimilar assessments). Additionally, Sussman and Gochyyev (2019) showed that the different dimensions captured patterns of children's strengths and weaknesses that would be ignored by a unidimensional version of the assessment. For instance, 11% of a preschool sample had a statistically significant discrepancy between their language and literacy score and their DRDP score for another dimension, with a practically significant median discrepancy equal to 6.7 months of development. The multidimensional approach provides greater opportunities for early detection of children's specific challenges, and ample additional evidence exists to support the current DRDP's multidimensional structure (Draney et al., 2022). Third, certain types, or strands, of validity evidence are salient across multiple purposes: the DRDP was aligned with the breadth and depth of content within California's early learning goals, including the Infant/Toddler Learning and Development Foundations, Preschool Learning Foundations, and Kindergarten Content Standards, including the Common Core State Standards, and the Next Generation Science Standards (Kriener-Althen et al., 2020). The reliability of DRDP's Language and Literacy Development scale is in the $r > .9$ range across different age levels, including with culturally and ethnically diverse samples (see Vogel et al., 2008). Invariance studies have found no differential item functioning (DIF) with DLLs, special education students, and multiple racial/ethnic groups (Draney et al., 2022). Again, a comparative analysis judged DRDP as having strong potential for culturally and linguistically responsive assessment (Ackerman, 2018b). Finally, it should be noted that validity concerns exist with all forms of child assessment including direct assessments (Waterman et al., 2012) and checklists (Bagnato et al., 2014). Readers may refer to NRC (2008) for a balanced discussion.

Data Analysis

This study modeled developmental trajectories, or growth curves, in language and literacy using a linear mixed-effects modeling framework. This is sometimes called multilevel modeling (Rabe-Hesketh & Skrondal, 2022) or hierarchical linear modeling (HLM). The outcome variable was each child's Language and Literacy Development score from the DRDP assessment, which is regarded as the most accurate and useful DRDP measure of language and early literacy with all children (Desired Results Collaborative Research Group, 2018). The multilevel modeling approach ignores measurement error in the outcome variable, but it is a common method in the developmental literature (e.g., Choi et al., 2018; Winsler et al., 2014), is a flexible framework for modeling longitudinal data with varying occasions and unequal spacing between occasions, and it produces consistent and unbiased estimates of the growth trajectories from latent dependent variables with heteroskedastic measurement error (Wang et al., 2019).

In terms of missing data, some of the variables had complete data (DLL status, DRDP score). Other variables had <0.5% of cases with missing data (race/ethnicity, sex). Analyses found no meaningful correlations between missing data and DRDP outcomes (e.g., the point-biserial correlation between missing race/ethnicity and DRDP score was 0.005). Several additional missing data-related concerns were explored: First, missing child-level data resulted in about 66,000 cases being dropped for an analytic n of 453,468 in the final model. The correlations between these missing data and the DRDP outcome were very small. Exploratory analyses compared models with and without the variables that caused most of the cases to be dropped from the models (special education status and child-care setting). The direction, magnitude, and significance of the effect estimates were comparable between models. Since the goal was to isolate the effect of home language, all variables were included in the final models. Second, selection bias mechanisms were considered but judged unlikely. DRDP is a state-mandated assessment, and the results generalize to a public ECE population that is required to complete the assessment. Third, models were tested using only the 109,107 children in the sample with three or more observations. The estimates were comparable with the full model and the trajectories were qualitatively identical. In sum, analyses found little evidence of consequential effects of missing data.

Multilevel Growth Curve Modeling

The analysis used multilevel modeling to estimate relationships between the dependent variable (each child's DRDP scaled score) and other variables. Supplement 4 in the online supplemental materials contains detailed information about the multilevel regression analyses including equations: This section provides an overview. The first research question asked how children progress through the DRDP developmental levels. To address this question, A three-level model (assessments, nested in children, nested in programs) was used to regress DRDP Language and Literacy Development scores on age, DLL status, sex, race/ethnicity, special education status, setting, and language match. The result produced a growth trajectory from birth through 60 months. Child age was centered at 3 years because this is when many children enter center-based ECE programs and there were benefits to interpreting some effects at this age (rather than at zero). The Level-3 variable was each child's program, not their teacher, because (a) many children in ECE programs have multiple teachers and fluid classrooms and (b) accounting for higher level variance supports accurate standard errors. The model included squared and cubic terms for age to relax the assumption of linear growth and give the trajectories flexibility. The results present three models of increasing complexity designed to address this research question. The α level was set to 0.001 for the three models due to the large sample sizes.

The second research question asked how home language moderates children's growth trajectories. To address this question, the DLL variable from Equation 1 was replaced with a set of dummy variables for 23 different home language categories. This parameterization set the baseline category to English (non-DLLs). Ethnicity/race variables were dropped because of multicollinearity (e.g., Spanish and Hispanic). Full interactions between home languages and the age terms were included to allow each groups' trajectories to conform to the data. As before, child age was centered at 3 years. For this research question, an α level of 0.05 was applied because disaggregation by home language and other variables produces some small cell sizes.

The Results section focused on interpreting growth trajectories and their variation across 23 different home language categories. However, only five home languages had minimally adequate ($n \geq 10$) numbers of children under 10 months old to support complete birth through kindergarten growth trajectories: Arabic, Cantonese, English (non-DLL), Mandarin (Putonghua), and Spanish. The other 18 languages were analyzed at other points on the age distribution, where adequate data existed. Some results were interpreted in terms of standardized effect sizes (*ES*) or months of typical child development: the denominator for the *ES* calculation was the within-age pooled standard deviation of the DRDP outcome. The denominator for the months of development calculation was the simple marginal linear growth per month in the full sample.

Transparency and Openness

The full sample contains all cases in a state database with valid DRDP scores (in the Language and Literacy Development domain, Comprehensive View) for years 2015–2018 that used the correct DRDP age level for the participant's age (e.g., we excluded cases where Infant/Toddler age level was used with children 3+ years old). The resulting dataset is the "full sample." We conducted analyses in Stata/MP 16.1 (StataCorp, 2019), including multilevel modeling with the *mixed, cov(un)* command. As this study is based on historical data, it was not preregistered and, presently, the data and analysis code are not publicly available.

Results

DLLs' and Non-DLLs' Language and Literacy Trajectories

Consistent with the focus of this paper, the results emphasize the interpretation of developmental trajectories and variation associated with home language. The first research question asked how children progress through the DRDP developmental levels. To address this question, each child's DRDP score was regressed on their age, DLL status, special education status, sex, race/ethnicity, childcare setting, and teacher–child language match. The results are interpretable as a model for expected growth, as a function of age, after controlling for the variables in the model (i.e., a developmental trajectory).

Table 4 contains the regression coefficients from three models that show steps in the model-building process. The regression coefficients, on a logit scale, are descriptive: They do not imply that certain factors cause less or greater child development. Model 1 is a three-level variance components model with random intercepts at Level 2 and Level 3. The intraclass coefficient (ICC) for Model 1 was 0.52 at Level 1 and 0.33 at Level 2. The values were above the conventional cutoff of 0.1 that indicates the need for multilevel modeling. Model 2 added fixed effects and the analytic sample size decreased to 843,115 observations (and a mean of 1.9 observations per child). The fixed effects include squared and cubic terms for child age to relax the assumptions of linear growth and allow the growth curve to take the functional form implied by the data. Age and Age^3 were significant in this model: Age^2 became nonsignificant when Age^3 was added, but a likelihood ratio (LR) test showed that the model with all three terms fit the data better than the model with only Age and Age^2, $\chi^2(1) = 15{,}160$, $p < 0.001$. All the binary fixed effects were significant and were retained. Each of the categorical variables (e.g., race/ethnicity) contributed significant contrasts and was retained. Model 3 added a random coefficient at Level 2, allowing each child to have their own trajectory. The LR test showed that Model 3 had better fit to the data than Model 2, $\chi^2(2) = 3{,}877$, $p < 0.001$, and was retained as the final model. The squared correlation between raw and predicted scores

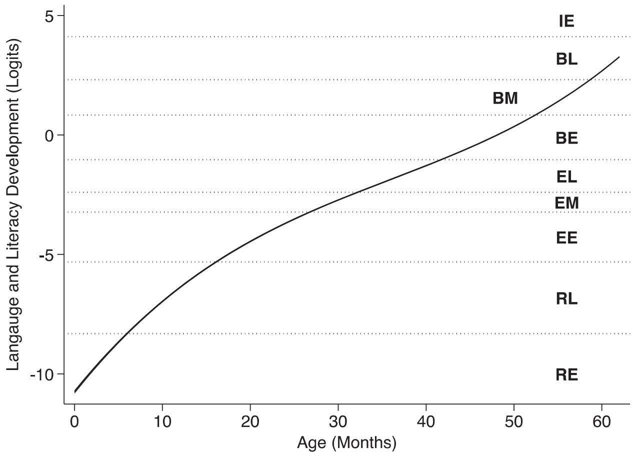
Table 4
Regression Coefficients for Children's Growth Trajectories

	Model 1			Model 2			Model 3		
	Coef.	*z*	*SE*	Coef.	*z*	*SE*	Coef.	*z*	*SE*
Fixed part									
Age				0.15***	319.4	−4.6E−04	0.14***	307.8	−4.6E−04
Age^2				2.9E−06	0.2	−1.3E−05	1.5E−05	1.2	−1.3E−05
Age^3				7.1E−05***	112.0	−6.4E−07	8.2E−05***	126.2	−6.5E−07
Dual language learner				−0.46***	−67.8	−0.01	−0.46***	−67.4	−0.01
Female				0.34***	72.0	−4.7E−03	0.34***	72.0	−4.7E−03
Black				−0.03**	−2.9	−0.01	−0.03	−2.5	−0.01
Hispanic				−0.14***	−17.5	−0.01	−0.14***	−17.3	−0.01
Asian/Pacific Islander				−0.03**	−3.1	−0.01	−0.04***	−3.8	−0.01
Other race/ethnicity				−0.11***	−8.0	−0.01	−0.11***	−7.9	−0.01
Special education				−0.72***	−75.8	−0.01	−0.72***	−75.6	−0.01
Language match				−0.11***	−17.0	−0.01	−0.11***	−16.5	−0.01
Family care				0.20***	5.7	−0.04	0.19***	5.6	−0.03
Head Start				0.07***	4.7	−0.01	0.06***	4.5	−0.01
Other settings				−0.05**	−2.9	−0.02	−0.05**	−2.9	−0.02
State school				−0.10***	−9.0	−0.01	−0.10***	−9.2	−0.01
Constant	0.49	19.5	−0.03	−1.56	−78.0	−0.02	−1.53	−76.5	−0.02
Random part									
SD (site)	1.67			0.89			0.89		
SD (child)	1.29			1.15			0.97		
SD (age)							0.03		
Correl (age, child)							0.10		
SD (residual)	2.01			1.36			1.34		
Log likelihood	−2,351,750			−1,647,038			−1,645,100		
N	1,035,370			843,115			843,115		

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 5

Children's Average Language and Literacy Growth Trajectory



Note. The 95% confidence interval is shown (visible in the left tail). RE = Responding Earlier; RL = Responding Later; EE = Exploring Earlier; EM = Exploring Middle; EL = Exploring Later; BE = Building Earlier; BM = Building Middle; BL = Building Later; IE = Integrating Earlier.

(i.e., R^2) was 0.50, indicating excellent fit. The small random coefficient for age (0.03) and the small correlation between the child-level random intercept and random slope (0.1) show that children who started lower tended to grow slightly less, and vice versa. The random intercepts standard deviations for site (0.89) and child (0.97) provide an idea of the variability around the mean growth trajectory.

Figure 5 contains a graph of the estimated average growth trajectory derived from Model 3, shown as a solid black line. The printed width of the line is often greater than that of its 95% confidence interval, which was graphed but only visible only in the left tail. The DRDP developmental levels are demarcated by horizontal dotted lines and labeled in the right column. Changes in the steepness of the line's slope are interpretable as changes in the rate at which children tend to progress through the logit scale, which, as mentioned, is also metered out in the (non-equidistant) levels. The trajectory shows an initial period of relatively rapid growth (0–10 months), followed by a decrease in the growth rate (10–30 months), a plateau (30–45 months), and then an acceleration (>45 months). Post hoc tests showed that these patterns were robust across DLL status and the age level-restricted datasets described in the method.

We can also interpret the trajectory in Figure 5 in terms of developmental levels. Development was more rapid during Responding Earlier (RE), slowed during Responding Later (RL) and EE, plateaued through Exploring Middle (EM) and Exploring Later (EL), and then accelerated again through the three Building levels (Building Earlier [BE], Building Middle [BM], and Building Later [BL]). Changes in a developmental trajectory are forward looking: a slope that changes toward the end of a developmental category is more related to the milestones in the subsequent developmental category than the current one. Those who wish to understand the developmental milestones that children tend to be working on when, for instance, development slows during the RL to EE transition, may review those levels of the learning progressions within the DRDP items—an illustrative example will be discussed later. The slope changes may be caused by the relative difficulty of the developmental milestones, maturation, environment, or some interaction thereof. Note that the three lowest levels (RE, RL, EE) are also the widest, indicating that progression through these stages requires a greater amount of development than later stages. As per convention with developmental trajectories, the current results reflect an estimated mean. In the real world, a population of children will have a distribution of developmental levels.

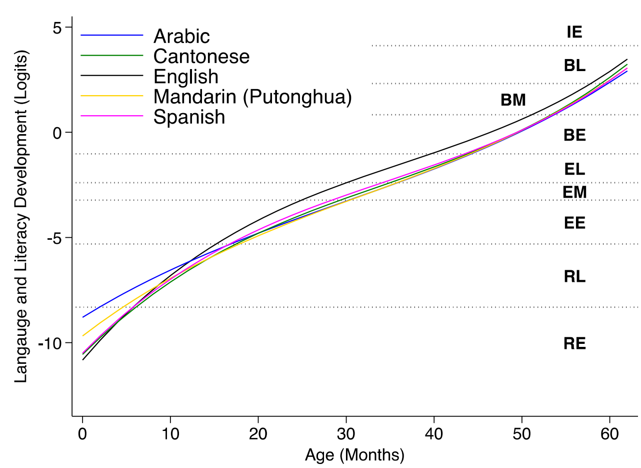
Moderating Effects of Home Language

The second research question asked how home language moderates children's growth trajectories. To address this research question, each child's DRDP score was regressed on their age, home language, sex, special education status, childcare setting, and teacher–child language match. Home language was a nominal variable with 23 levels, and English (non-DLL) was the reference category. Race/Ethnicity was removed because of its collinearity with language. For this analysis, the graph will be interpreted prior to the regression coefficients.

The graph in Figure 6 displays average estimated developmental trajectories for the five home languages that were judged to have adequate samples. Confidence intervals were omitted for clarity (see Supplement 5 in the online supplemental materials for graphs with confidence intervals). The trajectories are based on Model 4 (shown in Table 5) and are interpretable as the estimated differences attributable to home language when the other variables are held constant. The slopes for all five trajectories tended to decrease between 10 and 30 months and then increase after 45 months. In terms of between-group differences, on average, the groups of children classified as DLLs showed less development along the learning progression than their monolingual peers. For instance, at 3 years of age, the difference between English and Spanish-speaking DLLs was 0.59 logits ($ES = 0.30$), which translates into 3.6 months of

Figure 6

Language and Literacy Growth Trajectories for Five Home Languages



Note. RE = Responding Earlier; RL = Responding Later; EE = Exploring Earlier; EM = Exploring Middle; EL = Exploring Later; BE = Building Earlier; BM = Building Middle; BL = Building Later; IE = Integrating Earlier. See the online article for the color version of this figure.

Table 5

Regression Coefficients for Trajectories Disaggregated by Home Language

	Model 4		
	Coef.	*z*	*SE*
Fixed part			
Age	0.14***	244.0	5.7E−04
Age^2	−2.0E−04***	−11.9	1.7E−05
Age^3	8.5E−05***	99.2	8.6E−07
Arabic	−0.84***	−13.5	0.06
Armenian	−1.13***	−7.6	0.15
Cantonese	−0.71***	−15.3	0.05
Chaldean	−1.11***	−3.7	0.30
Farsi	−0.94***	−8.9	0.11
French	−0.30	−1.6	0.19
Hindi	−0.38*	−2.3	0.17
Hmong	−0.67***	−4.5	0.15
Japanese	−0.58**	−2.6	0.22
Khmer	−0.67***	−3.3	0.20
Korean	−0.57***	−5.4	0.10
Mandarin (Putonghua)	−0.83***	−10.4	0.08
Multilingual	−0.55***	−3.4	0.16
Other	−0.70***	−27.2	0.03
Pashto	−0.53	−1.8	0.30
Portuguese	−0.56***	−3.6	0.16
Punjabi	−1.41***	−12.6	0.11
Russian	−0.50***	−4.2	0.12
Spanish	−0.59***	−55.7	0.01
Tagalog	−0.92***	−6.6	0.14
Urdu	−0.77***	−4.1	0.19
Vietnamese	−1.27***	−17.9	0.07
Arabic × Age	0.01*	2.5	0.01
Armenian × Age	0.02	1.3	0.02
Cantonese × Age	0.01	1.7	4.0E−03
Chaldean × Age	0.02	1.2	0.02
Farsi × Age	0.01	0.6	0.01
French × Age	0.02	1.0	0.02
Hindi × Age	0.02	1.0	0.02
Hmong × Age	0.01	0.7	0.01
Japanese × Age	0.04*	2.1	0.02
Khmer × Age	−0.05	−2.0	0.03
Korean × Age	0.01	1.4	0.01
Mandarin (Putonghua) × Age	0.02*	2.2	0.01
Multilingual × Age	0.03	1.4	0.02
Other × Age	0.01***	5.6	2.2E−03
Pashto × Age	−0.02	−0.9	0.03
Portuguese × Age	−4.3E−03	−0.3	0.01
Punjabi × Age	0.05***	4.8	0.01
Russian × Age	−0.01	−1.3	0.01
Spanish × Age	8.0E−04	1.0	8.3E−04
Tagalog × Age	0.03**	2.6	0.01
Urdu × Age	−0.01	−0.8	0.02
Vietnamese × Age	0.03***	4.7	0.01
Arabic × Age^2	1.0E−03***	3.9	2.6E−04
Armenian × Age^2	2.5E−03	1.8	1.4E−03
Cantonese × Age^2	6.5E−04***	5.9	1.1E−04
Chaldean × Age^2	−2.2E−05	0.0	2.0E−03
Farsi × Age^2	1.1E−03*	2.5	4.4E−04
French × Age^2	2.5E−04	0.4	6.7E−04
Hindi × Age^2	1.0E−03	1.1	9.1E−04
Hmong × Age^2	−1.1E−04	−0.2	5.5E−04
Japanese × Age^2	9.1E−04	0.9	1.0E−03
Khmer × Age^2	0.01**	2.9	2.0E−03
Korean × Age^2	3.4E−04	0.9	3.8E−04
Mandarin (Putonghua) × Age^2	8.2E−04**	2.9	2.8E−04
Multilingual × Age^2	3.1E−04	0.2	1.6E−03
Other × Age^2	1.0E−03***	9.7	1.1E−04
Pashto × Age^2	−5.7E−05	−0.1	1.0E−03
Portuguese × Age^2	4.9E−04	1.0	4.9E−04
Punjabi × Age^2	1.6E−03**	3.1	5.2E−04
Russian × Age^2	4.2E−04	1.1	4.0E−04
Spanish × Age^2	4.4E−04***	15.7	2.8E−05
Tagalog × Age^2	1.6E−03*	2.4	6.6E−04
Urdu × Age^2	1.7E−03**	2.6	6.7E−04
Vietnamese × Age^2	2.5E−03***	6.0	4.1E−04
Arabic × Age^3	−4.4E−05***	−4.3	1.0E−05
Armenian × Age^3	−7.2E−05	−1.9	3.9E−05
Cantonese × Age^3	−8.0E−06	−1.3	6.1E−06
Chaldean × Age^3	2.3E−05	0.4	6.2E−05
Farsi × Age^3	−3.7E−05*	−2.2	1.7E−05
French × Age^3	−2.7E−05	−0.9	3.1E−05
Hindi × Age^3	−6.4E−05*	−2.0	3.2E−05
Hmong × Age^3	−4.2E−06	−0.2	1.9E−05
Japanese × Age^3	−7.5E−05	−1.9	4.0E−05
Khmer × Age^3	−1.2E−04*	−2.3	5.4E−05
Korean × Age^3	−2.9E−05	−1.7	1.7E−05
Mandarin (Putonghua) × Age^3	−3.2E−05*	−2.4	1.3E−05
Multilingual × Age^3	−3.9E−05	−0.9	4.2E−05
Other × Age^3	−3.1E−05***	−7.6	4.1E−06
Pashto × Age^3	6.9E−06	0.2	3.9E−05
Portuguese × Age^3	−8.1E−06	−0.3	2.4E−05
Punjabi × Age^3	−7.1E−05***	−4.3	1.7E−05
Russian × Age^3	−1.1E−06	−0.1	1.9E−05
Spanish × Age^3	−8.2E−06***	−6.1	1.3E−06
Tagalog × Age^3	−6.7E−05**	−3.3	2.1E−05
Urdu × Age^3	−2.6E−05	−1.0	2.6E−05
Vietnamese × Age^3	−7.5E−05***	−6.0	1.3E−05
Female	0.34***	71.9	4.7E−03
Special education	−0.72***	−75.5	0.01
Language match	−0.10***	−15.0	0.01
Family setting	0.19***	5.5	0.03
Head Start	0.06***	4.5	0.01
Other settings	−0.05	−2.9	0.02
State school	−0.10***	−9.3	0.01
Constant	−1.55	−80.1	0.02
Random part			
SD (site)	0.90		
SD (child)	0.98		
SD (age)	0.03		
Correl (age, child)	0.11		
SD (residual)	1.34		
N	843,349		

(*table continues*)

* $p < .05$. ** $p < .01$. *** $p < .001$.

typical development. The differences between English and the four groups of DLLs grew in early infancy, reached a maximum at around 30 months, and then narrowed as children approached school age. For all four DLL groups, the growth rates accelerated in earnest during early BE and, in middle BE, outpaced the growth for English monolinguals. Cantonese-speaking children made especially strong gains during the Building levels. A possible explanation for this "catch-up" effect is that the effects of opportunity gaps started to manifest around 1-year-old, and later, preschool participation helped children to narrow the opportunity gaps and catch up to their peers. The 0- to 10-month tail for the Arabic distribution is based on a small number of children and should be interpreted with caution.

These results offer partial support for a linguistic distance hypothesis: After controlling for the model covariates, children learning Spanish (with greater similarity to English) demonstrated greater language and literacy development over most of the distribution than those who spoke Cantonese, Arabic, or Mandarin (Putonghua). The differences between Spanish and others were statistically significant between 10 and 40 months, but after 40 months all the trajectories converged. Perhaps factors more influential than linguistic distance took hold as children approached school age. The *ES* of the differences between languages should be considered small. For instance, at 2 years, the *ES* of the difference between Spanish and Mandarin was 0.15, or 1.8 months of typical development; at 3 years, it was 0.11 or 1.3 months. In sum, although linguistic distance explained some variation in children's developmental trajectories, further investigation is needed to weigh the specific contribution of linguistic distance against other influences on development.

Additional comparisons between the trajectories for all 23 home language categories showed common trends across languages. These analyses focused on the 36- to 60-month age range where each of the languages had adequate data for growth modeling. For all language categories, children's rates of development accelerated approaching school age. The specific timing of the acceleration was slightly different between languages (see Supplement 5 in the online supplemental materials) and fell into three groups. Trajectories accelerated in late EL for four groups (Armenian, Khmer, Urdu, and Vietnamese), in early BE for 11 groups (Arabic, Cantonese, Farsi, Mandarin [Putonghua], Other, Pashto, Portuguese, Punjabi. Russian, Spanish, and Tagalog), and in middle BE for eight groups (Chaldean, English, French, Hindi, Hmong, Japanese, Korean, and Multilingual). Post hoc analyses, including correlations between timing and the research variables or qualitative groupings based on linguistic distance, did not explain group membership. Statistical variation may play some role in the observed differences. On balance, the similarities across languages outweigh the subtle differences, and the results provide evidence of common cross-linguistic patterns in language and literacy development per teacher reports using the DRDP assessment.

Table 5 contains the regression coefficients for Model 4. The model estimated 103 fixed effects, four variances for the random effects and one covariance between the Level-2 random intercept and random slope. The fit of Model 4 is roughly similar to that of Model 3, but the added parameters give us the ability to disaggregate Spanish-speaking DLLs (79% of the DLL sample) from other home languages. We can organize the fixed effects into three groups that reflect the mean differences in either intercepts or slopes across home languages. Moving from the top to bottom of Table 5, the first group contains the constant and coefficients for age, age^2, and age^3. The constant is the English (non-DLL) reference category's estimated development at 3 years of age. The second group of fixed effects contains the regression coefficients for (a) differences at the intercept across home language categories and (b) the effects of the other demographics variables. The home language coefficients are interpretable as differences at 3 years of age after controlling for other variables. The demographic variable coefficients are similarly interpretable as the vertical displacement of growth trajectories associated with group membership (e.g., female) after controlling for all other variables in the model—not necessarily the influence of a variable on children's rates of development. Being female is associated with higher language and literacy scores; being in special education is associated with lower scores. The coefficients for language match and most levels of the setting variable were also statistically significant. Teacher–child language match was associated with slightly lower scores. Being in a family daycare, or secondarily a Head Start program, was associated with slightly higher scores than center care (the reference category) or other settings, whereas being in a state school was associated with slightly lower scores than the other settings. These results could be related to various factors, including uncontrolled differences in the populations served in the different environments. Again, the coefficients should not be interpreted as differences in educational effectiveness, though this is one possible explanation. Additional research could examine differential growth rates associated with different environments. Post hoc analyses showed that the reported coefficient for language match reasonably generalizes across language categories: interaction coefficients between language category and language match showed a distribution of values, but the estimates for the larger groups typically hovered around the reported language match estimate.

The third and final group of fixed effects in Table 5 contains the regression coefficients for home language multiplied by age, age^2, and age^3. These interaction terms give each of the modeled trajectories the flexibility to conform to the empirical data. Statistically significant coefficients indicate that the curvature for the respective home language trajectory differs from the English reference category. These types of coefficients are generally difficult to interpret: the graphs in Supplement 5 in the online supplemental materials show the effects of these coefficients.

Discussion

Summary of Key Findings

This study examined the language and literacy development of linguistically diverse DLLs in California's public early education and kindergarten system that is responsible for providing equitable opportunities to learn for the state's large and growing population of multilingual children. We examined children's developmental trajectories from early infancy through kindergarten, interpreting the changes in growth rates over time and between groups of children who spoke different home languages. Three key findings emerged from this study. First, children's rate of development, after an early period of rapid growth from birth through 10 months, slowed between 10 and 30 months, plateaued between 30 and 45 months, and then accelerated though 60 months of age. Second, the trends were similar across different home languages with the following caveats: (a) DLLs' growth rates slowed more than their monolingual peers' between 10 and 30 months but grew more sharply after 45 months, and (b) the timing of the growth rate acceleration after 45 months differed slightly between home languages. Three, and contributing to an emerging area of research, a linguistic distance hypothesis was partially supported: when the effects of home language were isolated by the statistical models, Spanish-speaking children demonstrated greater levels of learning and development (over most of the age distribution) than four groups of children with linguistic structures less similar to English including Arabic, Cantonese, and Mandarin (Putonghua).

Interpreting Similarities and Differences Across Home Languages

Children who spoke different home languages had similar developmental trajectories relative to age (i.e., growth decelerated and then accelerated at similar ages). The results imply common developmental patterns and, in the context of other DRDP validity evidence (e.g., measurement invariance), support the cross-language applicability of the learning progressions. Indeed, prior research shows that DLLs' L2 (English) development, though often behind that of their monolingual peers', has similar early trajectories in areas of vocabulary, grammar, speech sound, and writing (Hammer et al., 2014). Longitudinal studies that modeled English vocabulary trajectories from preschool to kindergarten using nonlinear (quadratic) trajectories (Choi et al., 2018; Hoff & Ribot, 2017) also found these similarities between DLLs and their monolingual English-speaking peers. The late "catch-up" effect for DLLs found in the current study agrees with the extant literature showing that, although DLLs may perform below monolinguals during preschool years, they tend to catch up in late preschool and early elementary school in areas of speech sound production, vocabulary, and letter-word knowledge (Hammer et al., 2014). Hence, the current study extends the longitudinal literature in three ways: it (a) modeled development over a longer span, providing a continuous trajectory from birth through kindergarten; (b) included quadratic *and* cubic terms to achieve a more nuanced model of age-related changes; and (c) added a learning progression-based interpretive model that translated quantitative results into developmental milestones in ways that may be educationally useful, as described next.

Learning Progressions

Interpreting the estimated developmental trajectories in terms of the DRDP's learning progressions provides psychometrically accurate information about the milestones that are most salient when the growth rates change. Across home languages, children's growth rates tended to decrease during the RL level and the beginning of the next level, EE. For each DRDP item, we can examine the milestones associated with the RL to EE transition if we wish to better understand what children were working on as their growth rates declined. For instance, the expressive language item's summary definitions in Figure 1 show that, in RL, children are learning to use sounds or nonverbal strategies, and in EE children are using their first words to communicate. The results indicate that this transition was particularly difficult for children, regardless of home language, compared to other milestones in the learning progression. The RL to EE transition was also when differences between monolingual English speakers and groups of DLLs emerged, which suggests that this transition may be especially difficult for DLLs. Later, children's growth rates increased during BE and through the next two levels: BM and BL. Returning to the expressive language item, in these levels, children are learning to communicate in more complex phrases. Here, DLLs tended to grow faster than their monolingual English-speaking peers. The acceleration in growth rate may be the beginning of the catch-up effect documented in other research; caused by some combination of the task, environment, and internal development. An identical analysis can be repeated for all the DRDP items and any developmental level to create a developmental "profile." Readers will note that this sort of information goes far beyond that what is typically obtained from large-scale assessments, even compared to approaches that model trajectories through normative zones (e.g., Betebenner, 2009).

Interpreting Childcare Setting and Language Match Results

The significant coefficients for childcare setting and language match should be interpreted as small differences between groups (controlling for other potentially confounding factors), and not estimates of the effectiveness of a certain type of program or experience. First, family care and Head Start programs were assessed slightly higher than other settings, whereas state school programs were assessed slightly lower. These results may be due to systematic differences in the populations served in different settings rather than the effectiveness of a certain type of setting. Second, children classified as DLLs with a teacher who spoke their home language were assessed slightly lower than their matched peers with a teacher who did not speak their home language. This finding is inconsistent with the extensive evidence that bilingual education supports stronger child-level outcomes for DLLs. It is also inconsistent with recent evidence showing that language match, per se, led to fewer perceived barriers to home-school engagement among caregivers from culturally and linguistically diverse families (Li et al., 2021). However, Cook et al. (2023) showed that language match is a complex variable, as children are unlikely to be randomly assigned to a language match condition. In sum, additional research is needed to better understand the relationships between setting, language match, and child-level outcomes. Future studies should examine interactions between variables and within-group variation. If this complexity can be better understood, then studies can model the association between the variables and children's growth rates to examine questions about educational effectiveness.

Implications for Culturally and Linguistically Responsive Assessment

The current study is an important step toward validating the cross-linguistic applicability of the DRDP's learning progressions. The DRDP's item-level learning progressions can be viewed as interrelated models of development that offer the possibility of formulating testable hypotheses about similarities and differences in the sequence and rate of development in culturally and linguistically diverse children. Although this study supports the idea that, relatively speaking, children with different home languages experience similar levels of struggle and success with the DRDP's developmental levels in language and literacy, specific milestones in certain items may also contribute differentially to the differences between DLLs and their peers. A more granular focus, such as item-by-item examination of variation in the rate of children's development along individual learning progressions (i.e., a type of DIF by home language accounting for age) could examine the accuracy of the individual progressions and whether some groups tend to display specific developmental strengths to be leveraged and/or weakness to be supported. Prediction models based on large-scale data may help to disentangle characteristics of language disability or delay from typical development.

At the classroom level, teachers' pedagogically relevant understanding of child development is widely believed to be a cornerstone of effective practice with DLLs (Alvarez et al., 2014; Davison, 2019; Heritage et al., 2012). Teachers' use of observational assessments, including the DRDP, can support their understanding of typical sequences of learning and development (Jablon et al. 1999; NRC, 2008; Yun et al., 2021). When teachers understand learning along a continuum, they are more able to diagnose a child's current knowledge state and decide on the next teaching moves that will help the child to take incremental steps toward broader goals (Heritage, 2008). The current findings suggest that certain developmental transitions will be harder, or easier, for DLLs compared to their peers. Teachers who apply this understanding of the trajectories and the milestones at which they diverge may have greater ability to pinpoint children's difficulties, modify curriculum, interactions, or environments to target possible developmental "sticking points," and better understand the degree to which children are progressing as expected.

At the program and policy levels, aggregated results interpreted using the DRDP's learning progressions may support additional types of decision making. DRDP data are currently summarized for accountability and program improvement (Kriener-Althen et al., 2020), but real-world practice and its consequences are poorly understood. An administrator's analysis of the distribution of developmental levels in a group of children may support targeted modifications to curriculum, training, staffing, or other resources. At the level of the education system, the DRDP's learning progression is intended as an organizing framework for a large array of transactional structures in the classroom, curriculum, standards, assessment materials, and other processes and artifacts (Wilson, 2018). Evidence suggests that the DRDP's learning progressions are used effectively in California and Illinois to guide education systems (Yun et al., 2021). Thus, further investigation of the DRDP's cross-cultural validity—and its use at multiple levels of the education system—offers the potential for more culturally and linguistically responsive education for the hundreds of thousands of children assessed annually.

Limitations

Typically speaking, large-scale assessment datasets are rich sources of information, but they also present certain challenges related to sampling, data quality, and availability of information. First, selection bias is a perennial issue in most quantitative studies and is a concern here. This possibility was mitigated by the state-mandated nature of the DRDP and the fact that the analytic sample is approaching the size of the population of California children who attended public ECE during the study. However, the results may not generalize to populations of children who are under served or excluded by the public ECE system through limited access to care and/or structural racism. Second, there were very few children under 36 months of age for most home language groups. Generalizations about linguistic distance were based on the results from only five home languages with minimally sufficient numbers of young children to support a rigorous analysis. Relatedly, the small sample sizes for very young children did not allow a rigorous study of "early ECE attenders," and the degree to which children who attend public ECE from an early age have fundamentally different trajectories than those who enter at an older age. Ongoing data collection is generating larger samples of very young children to use in follow-up analyses. Third, children's demographic data were entered by their teachers. Human error should be expected, along with errors due to limitations in a teacher's knowledge about the child's family. Although relationships between missing data or errors and children's outcomes that induced spurious findings were possible, exploratory analyses and model comparisons with subsets of data were conducted and generated no evidence of problems. Fourth, our dataset does not contain potentially important child-level variables such as quantity and quality of exposure to L1 and L2, language onset and balance (including relative shifts in children's L1 and L2 proficiency as they grow), and family-level variables such as parental language proficiency, immigration status, and SES. Many (if not most) DLLs were likely assessed in English only. A bilingual assessment may have provided a more accurate picture of children's knowledge and skills. The way that a child balances their use of L1 and/or L2 during an assessment, for example, could correlate with both home language and teacher ratings and possibly influence the linguistic distance results. Also, the public preschool sample is SES-censored, but it is likely that a reliable indicator of SES would explain some variance. Considering these limitations, we also note that the current results, based on teachers' reports as structured by the DRDP's rubrics, are accurate and educationally meaningful descriptions of the target population's developmental trajectories. In the future, linking administrative and family data to the state assessment database would support deeper, more explanatory analyses of the relationships between basic sociolinguistic processes and children's developmental trajectories. Fifth, in terms of the DRDP assessment, reliability and rater effects are reasonable concerns for any observational measure. There is an expectation that score variance will be partly attributable to the child and partly to the rater. However, in early childhood, all forms of assessment have reliability challenges (Atkins-Burnett, 2007), the benefits of observational assessments have been well documented ((Moreno & Klute, 2011), and in this study, the large sample mitigates issues with unreliable assessment. In sum, future studies that address the above limitations should produce deeper understanding of language and literacy trajectories and help identify practices and policies that can better support the large and growing population of children classified as DLLs.

Conclusion

This study examined language and literacy trajectories among DLLs, who as a population continue to grow and in size and diversity, and their peers from monolingual English-speaking homes. In this large sample of children from public ECE and kindergarten settings, DLLs who spoke different home languages and monolingual English-speaking children demonstrated similar language and literacy trajectories. Children's rates of development, generally measured in English, slowed between 10 and 30 months, plateaued between 30 and 45 months, and then accelerated through 60 months of age. Between-group differences also existed: The growth rates for the groups of DLLs slowed more than their monolingual English-speaking peers between 10 and 30 months but then grew more sharply after 45 months. Certain developmental milestones along a learning progression were implicated in the early emergence of differences between DLLs and their monolingual peers, whereas other developmental milestones were salient when DLLs caught up to their monolingual peers. In addition, a linguistic distance hypothesis was partially supported: Greater similarity between a home language and English was associated with greater learning and development over much of the age distribution. The current study adds new, instructionally relevant, information about the similarities and differences in children's language and literacy trajectories among children from different language backgrounds.

References

Ackerman, D. J. (2018a). *Real world compromises: Policy and practice impacts of kindergarten entry assessment-related validity and reliability challenges* (Research Report No. RR-18-13). Educational Testing Service. https://doi.org/10.1002/ets2.12201

Ackerman, D. J. (2018b). *Comparing the potential utility of kindergarten entry assessments to provide evidence of English learners' knowledge and skills* (Research Report No. RR-18-36). Educational Testing Service. https://doi.org/10.1002/ets2.12224

Ackerman, D. J. (2019). Early childhood care and education workforce issues in implementing assessment policies. In C. P. Brown, M. B. McMullen, & N. File (Eds.), *The Wiley handbook of early childhood care and education* (pp. 537–566). Wiley-Blackwell. https://doi.org/10.1002/9781119148104.ch24

Aguirre-Muñoz, Z., & Amabisca, A. A. (2010). Defining opportunity to learn for English language learners: Linguistic and cultural dimensions of ELLs' instructional contexts. *Journal of Education for Students Placed at Risk*, *15*(3), 259–278. https://doi.org/10.1080/10824669.2010.495691

Akers, L., Grosso, P. D., Snell, E., Atkins-Burnett, S., Wasik, B. A., Carta, J., Boller, K., & Monahan, S. (2016). *Tailored teaching: Emerging themes from the literature on teachers' use of ongoing child assessment to individualize instruction*. Mathematica.

Alvarez, L., Ananda, S., Walqui, A., Sato, E., & Rabinowitz, S. (2014). *Focusing formative assessment on the needs of English language learners*. WestEd.

Atkins-Burnett, S. (2007). *Measuring children's progress from preschool through third grade*. Mathematica.

Bagnato, S. J., Goins, D. D., Pretti-Frontczak, K., & Neisworth, J. T. (2014). Authentic assessment as "best practice" for early childhood intervention: National consumer social validity research. *Topics in Early Childhood Special Education*, *34*(2), 116–127. https://doi.org/10.1177/0271121414523652

Banerjee, R., & Luckner, J. (2014). Training needs of early childhood professionals who work with children and families who are culturally and linguistically diverse. *Infants and Young Children*, *27*(1), 43–59. https://doi.org/10.1097/IYC.0000000000000000

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51. https://doi.org/10.1111/j.1745-3992.2009.00161.x

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Black, P., Wilson, M., & Yao, S.-Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives*, *9*(2–3), 71–123. https://doi.org/10.1080/15366367.2011.591654

Blom, E., Boerma, T., Bosma, E., Cornips, L., van den Heuij, K., & Timmermeister, M. (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, *40*(2), 151–171. https://doi.org/10.1177/0142723719892794

California Department of Education, Early Learning and Care Division and Special Education Division. (2015a). *DRDP (2015): A developmental continuum from early infancy to kindergarten entry: Infant/toddler view for use with infants and toddlers*. https://www.cde.ca.gov/sp/cd/ci/documents/drdp2015infanttoddler.pdf.

California Department of Education, Early Learning and Care Division and Special Education Division. (2015b). *DRDP (2015): An early childhood developmental continuum: Preschool comprehensive view*. https://www.desiredresults.us/sites/default/files/docs/forms/DRDP2015_PSC_Comprehensive_View_Combined-20200219_ADA.pdf.

Cannady, K., Jackle, E., & Hilegass, M. (2019). *Mastering KIDS: Voices from the field*. Teach Plus. https://teachplus.org/sites/default/files/publication/pdf/mastering_kids_til.pdf

Castro, D. C. (2014). The development and early care and education of dual language learners: Examining the state of knowledge. *Early Childhood Research Quarterly*, *29*(4), 693–698. https://doi.org/10.1016/j.ecresq.2014.08.003

Castro, D. C., & Artiles, A. J. (2021). *Language, learning, and disability in the education of young bilingual children.* Multilingual Matters.

Castro, D. C., Espinosa, L., & Páez, M. (2011). Defining and measuring quality in early childhood practices that promote dual language learners' development and learning. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 257–280). Paul H. Brookes.

Choi, J. Y., Rouse, H., & Ryu, D. (2018). Academic development of Head Start children: Role of dual language learning status. *Journal of Applied Developmental Psychology*, *56*, 52–66. https://doi.org/10.1016/j.appdev.2018.02.001

Collier, V. P., & Thomas, W. P. (2017). Validating the power of bilingual schooling: Thirty-two years of large-scale, longitudinal research. *Annual Review of Applied Linguistics*, *37*, 203–217. https://doi.org/10.1017/S0267190517000034

Conboy, B. (2013). *Neuroscience research: How experience with one or multiple languages affects the developing brain. California's best practices for young dual language learners: Research overview papers*. California Department of Education.

Cook, K. D., Ferreira van Leer, K., & Lombardi, C. M. (2023). Exploring predictors and outcomes of racial/ethnic match between children and providers in Early Head Start. *Early Childhood Research Quarterly*, *62*, 275–286. https://doi.org/10.1016/j.ecresq.2022.09.002

Davison, C. (2019). Using assessment to enhance learning in English language education. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 433–454). Springer. https://doi.org/10.1007/978-3-030-02899-2_21

Department of Health and Human Services (DHHS). (2008). *Dual language learning: What does it take? Head Start dual language report*. Administration for Children and Families.

Desired Results Collaborative Research Group. (2018). *Technical report for the Desired Results Developmental Profile (2015)*. California Department of Education.

Diakow, R., Irribarra, D. T., & Wilson, M. (2013). Some comments on representing construct levels in psychometric models. In R. E. Millsap, L. A. van der Ark, D. M. Bolt, & C. M. Woods (Eds.), *New developments in quantitative psychology* (pp. 319–334). Springer. https://doi.org/10.1007/978-1-4614-9348-8_20

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*(4), 305–310. https://doi.org/10.3102/0013189X10370204

Draney, K., Sussman, J., Gochyyev, P., Kriener-Althen, K., Newton, E. K., & Mangione, P. (2022). *DRDP technical report for early infancy through kindergarten: Structural validity and reliability for the desired results developmental profile*. California Department of Education. https://www.desiredresults.us/sites/default/files/docs/resources/research/DRDP%20early%20infancy%20thru%20K%20tech%20report%2020211216_FINAL%20ADA_V3.pdf

Draney, K., & Wilson, M. (2011). Selecting cut scores with a composite of item types: The construct mapping procedure. *Journal of Applied Measurement*, *12*(3), 298–309.

Dunphy, E. (2010). Assessing early learning through formative assessment: Key issues and considerations. *Irish Educational Studies*, *29*(1), 41–56. https://doi.org/10.1080/03323310903522685

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123–182. https://doi.org/10.1080/03057267.2011.604476

Escobar, K., & Tamis-Lemonda, C. (2015). Understanding the variability in Latino children's dual-language development. In N. J. Cabrera, B.

Leyendecker, & R. Bradley (Eds.), *Positive development of minority children* (pp. 89–106). Springer.

Espinosa, L. M., & López, M. L. (2007). *Assessment considerations for young English language learners across different levels of accountability*. First 5 LA.

Floccia, C., Sambrook, T. D., Luche, C. D., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot–Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*(1), 1–136.

Fortuny, K., Hernandez, D. J., & Chaudry, A. (2010). *Young children of immigrants: The leading edge of America's future*. The Urban Institute. https://doi.org/10.1037/e721852011-001

Gaitan, C. D. (2012). Culture, literacy, and power in family–community–school–relationships. *Theory Into Practice*, *51*(4), 305–311. https://doi.org/10.1080/00405841.2012.726060

Gardner-Neblett, N., & Iruka, I. U. (2015). Oral narrative skills: Explaining the language-emergent literacy link by race/ethnicity and SES. *Developmental Psychology*, *51*(7), 889–904. https://doi.org/10.1037/a0039274

Golinkoff, R. M., Deniz Can, D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gummer, E., & Mandinach, E. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, *117*(4), 1–22. https://doi.org/10.1177/016146811511700401

Hammer, C. S., Burchinal, M., Hong, S. S., LaForett, D. R., Páez, M., Buysse, V., Espinosa, L., Castro, D., & López, L. M. (2020). Change in language and literacy knowledge for Spanish–English dual language learners at school-entry: Analyses from three studies. *Early Childhood Research Quarterly*, *51*, 81–92. https://doi.org/10.1016/j.ecresq.2019.07.001

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D., & 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

Heckman, J. J., & Masterov, D. V. (2007). *The productivity argument for investing in young children*. National Bureau of Economic Research. https://www.nber.org/papers/w13016

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. The Council of Chief State School Officers. https://www.michiganassessmentconsortium.org/wp-content/uploads/Learning-Progressions.pdf

Heritage, M., Walqui, A., & Linquanti, R. (2012, August). *Formative assessment as contingent teaching and learning: Perspectives on assessment as and for language learning in the content areas* [Paper presentation]. Understanding Language Conference, Stanford, CA.

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., & Ribot, K. M. (2017). Language growth in English monolingual and Spanish–English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics*, *190*, 241–245.e1. https://doi.org/10.1016/j.jpeds.2017.06.071

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, *29*(4), 433–444. https://doi.org/10.1016/j.ecresq.2014.04.012

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Illinois State Board of Education. (2020). *2019–2020 Illinois Kindergarten Individual Development Survey (KIDS) report*.

Jablon, J. R., Dombro, A. L., & Dichtelmiller, M. L. (1999). *The power of observation*. Teaching Strategies.

Jackson, C. W., Schatschneider, C., & Leacox, L. (2014). Longitudinal analysis of receptive vocabulary growth in young Spanish English-speaking children from migrant families. *Language, Speech, and Hearing Services in Schools*, *45*(1), 40–51. https://doi.org/10.1044/2013_LSHSS-12-0104

Jacoby, J. W., & Lesaux, N. K. (2017). Language and literacy instruction in preschool classes that serve Latino dual language learners. *Early Childhood Research Quarterly*, *40*, 77–86. https://doi.org/10.1016/j.ecresq.2016.10.001

Kaiser, A., Eppenberger, L. S., Smieskova, R., Borgwardt, S., Kuenzli, E., Radue, E.-W., Nitsch, C., & Bendfeldt, K. (2015). Age of second language acquisition in multilinguals has an impact on gray matter volume in language-associated brain areas. *Frontiers in Psychology*, *6*, Article 638. https://doi.org/10.3389/fpsyg.2015.00638

Karelitz, T. M., Parrish, D. M., Yamada, H., & Wilson, M. (2010). Articulating assessments across childhood: The cross-age validity of the Desired Results Developmental Profile—Revised. *Educational Assessment*, *15*(1), 1–26. https://doi.org/10.1080/10627191003673208

Khalegi, F. (2010). *Educators' perceptions of the effectiveness of the desired results system: A study of the California department of education accountability initiative* [Unpublished doctoral dissertation]. California State University.

Kriener-Althen, K., Newton, E. K., Draney, K., & Mangione, P. L. (2020). Measuring readiness for kindergarten using the Desired Results Developmental Profile. *Early Education and Development*, *31*(5), 739–763. https://doi.org/10.1080/10409289.2020.1743160

Li, L.-W., Ochoa, W., McWayne, C. M., Priebe Rocha, L., & Hyun, S. (2021). "Talk to me": Parent–teacher background similarity, communication quality, and barriers to school-based engagement among ethnoculturally diverse Head Start families. *Cultural Diversity and Ethnic Minority Psychology*. Advance online publication. 1–12. https://doi.org/10.1037/cdp0000497

MacLeod, A. A. N., Castellanos-Ryan, N., Parent, S., Jacques, S., & Séguin, J. R. (2019). Modelling vocabulary development among multilingual children prior to and following the transition to school entry. *International Journal of Bilingual Education and Bilingualism*, *22*(4), 473–492. https://doi.org/10.1080/13670050.2016.1269718

Mancilla-Martinez, J., Hwang, J. K., Oh, M. H., & McClain, J. B. (2020). Early elementary grade dual language learners from Spanish-speaking homes struggling with English reading comprehension: The dormant role of language skills. *Journal of Educational Psychology*, *112*(5), 880–894. https://doi.org/10.1037/edu0000402

Mancilla-Martinez, J., & Lesaux, N. K. (2011). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development*, *82*(5), 1544–1560. https://doi.org/10.1111/j.1467-8624.2011.01633.x

Mangione, P. L., Osborne, T., & Mendenhall, H. (2019). What's next? How learning progressions help teachers support children's development and learning. *Young Children*, *74*(3), 20–25.

Markowitz, A. J., Bassok, D., & Grissom, J. A. (2020). Teacher-child racial/ethnic match and parental engagement with head start. *American Educational Research Journal*, *57*(5), 2132–2174. https://doi.org/10.3102/0002831219899356

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

McCarthy, E., Morgan, C., & Sussman, J. (2020). *Growing strong learners in Galt Joint Union Elementary School District*. WestEd. https://www.wested.org/resources/growing-strong-learners-in-galt-joint-union-elementary-school-district/#

Montoya-Ávila, A., Ghebreab, N., & Galindo, C. (2018). Toward improving the educational opportunities for Black and Latinx young children: Strengthening family-school partnerships. In S. Sonnenschein, & B. E.

Sawyer (Eds.), *Academic socialization of young Black and Latino children: Building on family strengths* (pp. 209–231). Springer.

Moreno, A. J., & Klute, M. M. (2011). Infant–toddler teachers can successfully employ authentic assessment: The Learning Through Relating system. *Early Childhood Research Quarterly*, *26*(4), 484–496. https://doi.org/10.1016/j.ecresq.2011.02.005

Murphey, D., Guzman, L., & Torres, A. (2014). *America's Hispanic children*. Child Trends.

National Academies of Sciences, Engineering, and Medicine. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. The National Academies Press.

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

National Education for the Association of Young Children. (2020). *Developmentally appropriate practice (DAP) position statement*.

National Education Goals Panel. (1995). *The National Education Goals Report: Building a nation of learners 1995*. National Education Goals Panel, U.S. Government Printing Office.

National Research Council. (2008). *Early childhood assessment: Why, what, and how*. The National Academies Press. https://doi.org/10.17226/12446

Neuman, S. B., & Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighborhoods. *Reading Research Quarterly*, *36*(1), 8–26. https://doi.org/10.1598/RRQ.36.1.1

Ntuli, E., Nyarambi, A., & Traore, M. (2014). Assessment in early childhood education: Threats and challenges to effective assessment of immigrant children. *Journal of Research in Special Educational Needs*, *14*(4), 221–228. https://doi.org/10.1111/j.1471-3802.2012.01256.x

Páez, M. M., Tabors, P. O., & López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *Journal of Applied Developmental Psychology*, *28*(2), 85–102. https://doi.org/10.1016/j.appdev.2006.12.007

Paradis, J., & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, *20*(1), Article e12433. https://doi.org/10.1111/desc.12433

Park, M., O'Toole, A., & Katsiaficas, C. (2017). *Dual language learners: A demographic and policy profile for California*. Migration Policy Institute.

Portes, A., & Rumbaut, R. G. (2014). *Immigrant America: A portrait*. UC Press. https://doi.org/10.1525/9780520959156

Rabe-Hesketh, S., & Skrondal, A. (2022). Multilevel and longitudinal modeling using Stata: Vol. I. *Continuous responses* (4th ed.). Stata Press.

Radman, N., Jost, L., Dorood, S., Mancini, C., & Annoni, J.-M. (2021). Language distance modulates cognitive control in bilinguals. *Scientific Reports*, *11*(1), Article 24131. https://doi.org/10.1038/s41598-021-02973-x

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development*, *89*(3), 929–940. https://doi.org/10.1111/cdev.12770

Rinaldi, C., & Páez, M. (2008). Preschool matters: Predicting reading difficulties for Spanish-speaking bilingual students in first grade. *Learning Disabilities*, *6*(1), 71–84.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Saxe, G. B. (1988). Candy selling and math learning. *Educational Researcher*, *17*(6), 14–21. https://doi.org/10.3102/0013189X017006014

Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, *103*, Article 101602. https://doi.org/10.1016/j.ijer.2020.101602

Schneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686. https://doi.org/10.1017/S0305000912000141

StataCorp. (2019). *Stata Statistical Software: Release 16* [Computer software].

Stiggins, R. (2008). *Assessment manifesto*. Educational Testing Service.

Stipek, D. (2019). *Getting down to facts II—Technical report: Early childhood education in California*. Policy Analysis for California Education (PACE).

Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, *22*(2), 401–415. https://doi.org/10.1017/S1366728917000682

Sussman, J., & Gochyyev, P. (2019, April). *Multidimensional IRT analysis to support formative, longitudinal assessment of early childhood development*. In K. Draney (Chair), Supporting children's early learning and kindergarten goals with a standards-aligned assessment: Effective practices in statewide aggregate reporting [Paper presentation]. Annual Meeting of the American Educational Research Association, Toronto, Ontario, Canada. https://tinyurl.com/3a7uf6f9

Sussman, J., Gochyyev, P., & Draney, K. (2021, April). *Examining the consequential validity of performance levels for a kindergarten entry assessment (KEA): Item-level consistency and grain size*. Paper presented at the annual meeting of the American Educational Research Association (online).

Sussman, J., Melnick, H., Newton, E., Kriener-Althen, K., Draney, K., Mangione, P., & Gochyyev, P. (2022). *Preschool quality and child development: How are learning gains related to program ratings?* Learning Policy Institute. https://learningpolicyinstitute.org/product/california-preschool-quality-ratings-report

Tabors, P. O., Snow, C. E., & Dickinson, D. K. (2001). Homes and schools together: Supporting language and literacy development. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language* (pp. 313–334). Paul H. Brookes.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445. https://doi.org/10.1177/1367006911403202

U.S. Department of Health and Human Services & U.S. Department of Education. (2016). *Policy statement on supporting the development of children who are dual language learners in early childhood programs*.

Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Child Development*, *80*(5), 1545–1563. https://doi.org/10.1111/j.1467-8624.2009.01350.x

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis*, *37*(4), 612–637. https://doi.org/10.3102/0162373715573310

Vogel, C., Aikens, N., Atkins-Burnett, S., Martin, E. S., Caspe, M., Sprachman, S., & Love, J. M. (2008). *The First 5 LA universal preschool child outcomes study spring 2007 pilot study*. Mathematica.

Wang, C., Xu, G., & Zhang, X. (2019). Correction for item response theory latent trait measurement error in linear mixed effects models. *Psychometrika*, *84*(3), 673–700. https://doi.org/10.1007/s11336-019-09672-7

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly*, *27*(1), 46–54. https://doi.org/10.1016/j.ecresq.2011.06.003

Weisenfeld, G. G., Garver, K., & Hodges, K. (2020). Federal and state efforts in the implementation of kindergarten entry assessments (2011–2018). *Early Education and Development*, *31*(5), 632–652. https://doi.org/10.1080/10409289.2020.1720481

WestEd. (2022a). *Desired Results Developmental Profile (2015) early infancy through kindergarten: Interrater reliability*. California Department of Education.

WestEd. (2022b). *Desired Results Developmental Profile (2015) early infancy through kindergarten: Validity in relation to external assessments of child development*. California Department of Education.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716–730. https://doi.org/10.1002/tea.20318

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, *37*(1), 5–20. https://doi.org/10.1111/emip.12188

Wilson, M. (in press). *Constructing measures: An item response modeling approach* (2nd ed.). Routledge.

Wilson, M., & Lehrer, R. (2021). Improving learning: Using a learning progression to coordinate instruction and assessment. *Frontiers in Education*, *6*, Article 654212. https://doi.org/10.3389/feduc.2021.654212

Winsler, A., Burchinal, M. R., Tien, H. C., Peisner-Feinberg, E., Espinosa, L., Castro, D. C., LaForett, D. R., Kim, Y. K., & De Feyter, J. (2014). Early development among dual language learners: The roles of language use at home, maternal immigration, country of origin, and socio-demographic variables. *Early Childhood Research Quarterly*, *29*(4), 750–764. https://doi.org/10.1016/j.ecresq.2014.02.008

Wong-Fillmore, L., Ammon, P., McLaughlin, B., & Ammon, M. S. (1985). *Learning English through bilingual instruction*. National Institute of Education.

Yazejian, N., Bryant, D., Freel, K., & Burchinal, M. (2015). High-quality early education: Age of entry and time in care differences in student outcomes for English-only and dual language learners. *Early Childhood Research Quarterly*, *32*, 23–39. https://doi.org/10.1016/j.ecresq.2015.02.002

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development.

Yow, W. Q., & Markman, E. M. (2011). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development*, *12*(1), 12–31. https://doi.org/10.1080/15248372.2011.539524

Yun, C., Melnick, H., & Wechsler, M. (2021). *High-quality early childhood assessment: Learning from states' use of kindergarten entry assessments*. Learning Policy Institute. https://learningpolicyinstitute.org/product/high-quality-kindergarten-entry-assessments-report

Received November 15, 2021
Revision received November 30, 2022
Accepted December 6, 2022 ■