WORKING PAPER · NO. 2025-135

Introducing The SPEAK: A Scalable Computer-Adaptive Tool to Measure Knowledge of Early Human Development

Caroline Gaudreau, Dani Levine, John A. List, and Dana Suskind

OCTOBER 2025

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Introducing The SPEAK:
A Scalable Computer-Adaptive Tool to Measure Knowledge of Early Human Development

Caroline Gaudreau, Dani Levine, John A. List, Dana Suskind*
University of Chicago, ANU, and NBER

*Corresponding author: Dana Suskind, Department of Surgery, University of Chicago Medicine, 5841 S. Maryland Avenue, Chicago, IL 60637. Email: dsuskind@uchicago.edu

September 2025

Abstract

Research shows responsive caregiving enhances children's brain development, with parental knowledge predicting positive behaviors and outcomes. However, knowledge varies widely across educational levels, highlighting the need for targeted interventions. Despite evidence that this knowledge can be improved, no comprehensive metric exists for efficient assessment. We introduce SPEAK (Survey of Parent/Provider Expectations and Knowledge), a computer-adaptive tool grounded in item-response theory that we created, to address this gap by measuring parental and educator knowledge across development domains with precision and speed. This paper details SPEAK's development, including domain construction, cognitive interviewing, expert review, psychometric calibration, and validity evidence. SPEAK offers a flexible, scalable solution for clinical, educational, research, and policy settings. By identifying knowledge gaps, it enables tailored interventions, supports professional development, and informs policy, ultimately improving parent-child interactions and child outcomes. Our tool bridges critical gaps in assessing child development knowledge, advancing research and cross-sector collaboration to promote early childhood development worldwide.

Keywords: Early Education, Metrics, SPEAK, Measurement, Child Care, Parents, Educators, Field Experiments
JEL: C81 C93 I12 I18 I21 J13 O15

Acknowledgements
Thanks to our TMW team for helping us in the design phase of this project. Julie Pernaudet, Rebecca Royer, Liz Sablich, and Heidi Stevens provided useful comments on an earlier version of this study.

Introduction

Decades of research demonstrate that responsive, sensitive caregiving optimizes children's brain development across multiple domains, including language, cognitive, and socioemotional skills (Hirsh-Pasek et al., 2015a; Leung et al., 2020; Tamis-LeMonda et al., 2014). Practices such as engaging in back-and-forth conversations, responding to infants' cues, and reading books early foster optimal development (Romeo et al., 2018; Planalp et al., 2019; Dunst et al., 2012). Parental knowledge of child development significantly influences these caregiving behaviors, predicting improved child outcomes (Leung & Suskind, 2020; List et al., 2021). Yet, caregiver knowledge varies widely, particularly by educational attainment, and misconceptions persist (Bornstein et al., 2010; Rowe, 2008). Interventions can bridge this gap, with studies showing that teaching parents enhances interactions and child skills (List et al., 2021; Suskind et al., 2016).

Despite this evidence, no comprehensive, efficient tool exists to assess child development knowledge, hindering clinical, educational, research, and policy efforts. In clinical settings, pediatricians lack quick methods to identify knowledge gaps during brief well-visits, where discussions often prioritize physical over developmental topics (Halfon et al., 2011; Leung & Suskind, 2020). In education, the absence of standardized metrics limits tailored professional development for teachers, impacting child outcomes (Nocita et al., 2020). Researchers and policymakers also need reliable measures to track knowledge changes and inform interventions.

This study introduces a new tool to tackle this need: SPEAK (Survey of Parent/Provider Expectations and Knowledge). The SPEAK is a computer adaptive tool (CAT) that we developed. SPEAK leverages technology to measure parent and practitioner knowledge of multiple domains of child development in a fast, precise way. In the CAT, there is a large item bank of hundreds of items. An item is randomly selected from this item bank for the first administration. After, additional items are administered adaptively based on an item response theory (IRT)-based algorithm (derived from an initial calibration study): participants who answer correctly are presented with more difficult items, while participants who answer incorrectly receive progressively easier items. This process continues until the model converges on a precise score.

A CAT offers several distinct advantages over traditional fixed-length assessments. Higher precision is achieved because the adaptive algorithm selects items that provide maximum information about each participant's ability level. Rather than administering the same items to all participants regardless of their knowledge, the CAT targets the optimal difficulty level for each individual, resulting in more accurate ability estimates with fewer items. Broader content coverage becomes possible because the large item bank (hundreds of items) can encompass multiple domains and difficulty levels, whereas fixed-length tests are constrained by practical time limits to include only a subset of possible content areas.

Elimination of test-retest bias occurs because participants receive different sets of items in subsequent administrations, drawn from the same calibrated item bank. This prevents practice effects and item memorization that can inflate scores in traditional assessments. The faster testing experience results from the adaptive termination criteria—once the algorithm achieves the desired level of measurement precision, testing stops automatically. Participants typically complete 20-30 items rather than the 50-100 items often required in fixed-length assessments covering equivalent content.

The personalized experience emerges naturally from the adaptive algorithm: participants who demonstrate higher knowledge levels encounter more challenging items, while those with more knowledge gaps receive appropriately easier items. This reduces frustration for struggling participants and prevents boredom among high-performers. Finally, dynamic updating capability allows researchers to add new items to the bank as child development research evolves, without requiring re-calibration of the entire assessment. New items can be field-tested and integrated seamlessly, ensuring the tool remains current with emerging scientific knowledge.

The development of the SPEAK is inspired by the success of its predecessor, the SPEAK, a fixed-length tool demonstrated to be a valid and reliable measure of parental knowledge of child development (Suskind et al., 2018). Higher SPEAK scores have been associated with improved parental behaviors and child outcomes (e.g., Leung & Suskind, 2020; List et al., 2021). For instance, a randomized controlled trial by List et al. (2021) showed that interventions boosting SPEAK scores enhanced parent-child interactions, leading to gains in children's language, math, and social-emotional skills. Widely used in clinical and research settings, the SPEAK has demonstrated consistent effectiveness. The SPEAK builds on this foundation by offering broader content coverage, greater measurement precision, and improved efficiency.

We conclude that SPEAK has the potential to enhance considerably how we assess knowledge of child development, offering a fast, precise, and scalable solution for parents, educators, clinicians, policymakers, and researchers. By identifying knowledge gaps across critical domains, it empowers targeted interventions, enhances professional development, and informs evidence-based policy decisions. Its computer-adaptive design ensures efficiency without sacrificing depth, bridging the gap between research and real-world applications.

Ultimately, for academics, traditional assessments often rely on small-scale studies but SPEAK's computer-adaptive design allows academics to conduct research across larger, more representative populations, enhancing external validity and cross-disciplinary applications. In this manner, SPEAK addresses a critical methodological gap in developmental research by providing the first scalable, computer-adaptive tool for measuring parental and educator knowledge across multiple domains of child development. The assessment's efficiency—requiring only about 5 minutes to complete—combined with its precision enables researchers to conduct large-scale studies across diverse populations while minimizing participant burden.

This represents a significant advancement over traditional fixed-length measures that often limit researchers to smaller convenience samples. Additionally, the tool's adaptive design allows researchers to investigate knowledge as both a mediating variable and outcome in intervention studies, while the item bank's capacity for continuous updating ensures the assessment evolves with emerging scientific discoveries, creating a dynamic platform for advancing early childhood development research.

For policymakers, SPEAK provides actionable data for designing targeted early childhood interventions at scale (Al-Ubaydli et al., 2017). Current policy initiatives often rely on broad demographic proxies to identify families needing support, but these approaches miss substantial variation within socioeconomic groups. By efficiently measuring actual knowledge levels across domains like language development and socioemotional learning, policymakers can identify specific knowledge gaps that interventions should address. This precision targeting maximizes return on investment for limited public resources.

The tool also enables evidence-based evaluation of policy impacts in ways previously impossible. Policymakers can measure whether professional development programs for early childhood educators actually improve their developmental knowledge, or whether public health campaigns successfully increase parental understanding of critical concepts like responsive caregiving. The brief administration time makes population-level surveillance feasible, allowing jurisdictions to track knowledge trends and disparities over time. This data infrastructure supports more sophisticated policy development, moving beyond one-size-fits-all approaches toward interventions tailored to the specific knowledge profiles of different communities.

The remainder of our study proceeds as follows. In the next section we detail the development of the SPEAK. Section II describes its eight domains and their construction. Section III reviews the cognitive interviewing process, ensuring items are clear and accessible to the general population. Section IV presents expert feedback on content validity. Section V explains the calibration process. Section VI reports the validation study, providing evidence of the tool's convergent and divergent validity. Section VII proposes directions for future research. Section VIII offers concluding remarks.

II. Development of the SPEAK

Content Development: Domains of the SPEAK

The SPEAK organizes content into eight domains of early childhood knowledge, selected for their expected influence on caregiver behaviors and child outcomes. Five domains align with key school readiness indicators: language, literacy, math, cognitive, and socioemotional development (Pace et al., 2019). Three additional domains—screen media learning, dual language learning, and brain development—are included, as detailed below.

Language development encompasses children's ability to understand and use language for communication. This domain is critical, as early language skills predict not only sustained language proficiency through adolescence (Rescorla, 2011) but also enhanced social skills (e.g., reduced externalizing and internalizing behaviors; Hentges et al., 2021), academic achievement in reading and math (Pace et al., 2019), and higher occupational status in adulthood (Johnson et al., 2010). Caregivers significantly influence children's language development.

The literature shows that the quantity of language exposure matters: diverse adult vocabulary supports skill growth (Rowe, 2012; Golinkoff et al., 2019). Equally important is the quality of linguistic input. For instance, infants learn words more effectively through infant-directed speech, characterized by high-pitched, playful tones and exaggerated sounds (Golinkoff et al., 2015; Thiessen et al., 2005). Back-and-forth conversations also foster language acquisition (Hirsh-Pasek et al., 2015a; Romeo et al., 2018). Interventions can enhance these skills by teaching caregivers strategies aligned with developmental science (List et al., 2021). Notably, List et al. (2021) found that increasing parental knowledge of child development, measured by the SPEAK tool, improved caregiver-child interactions, leading to higher child language scores.

Literacy development refers to the children's reading and writing skills. Early literacy proficiency strongly predicts later outcomes: for instance, oral narrative skills at age four correlate with reading comprehension in high school (Suggate et al., 2018), and kindergarten reading skills predict reading and math achievement beyond fifth grade (Rabiner et al., 2016). Caregivers foster early literacy in part by reading regularly to children (Demir-Lira et al., 2019).

Engaging practices, such as asking questions during stories and relating narratives to children's lives, enhance these benefits, improving print concept knowledge (Sim et al., 2013). Caregiver beliefs about early literacy's importance are linked to preschoolers' emerging skills (Bennett et al., 2002), suggesting these beliefs shape literacy-promoting behaviors. Similarly, preschool teachers' literacy beliefs influence classroom practices, with brief professional development interventions aligning both with current research (Lynch, 2017).

Math development encompasses children's skills in areas such as number knowledge and spatial competence. Early math abilities correlate with other developmental domains, including language and literacy (Purpura et al., 2017), and kindergarten math skills predict academic achievement in math, reading, and science by eighth grade (Claessens & Engel, 2013).

Caregivers can foster math learning before school entry through math and spatial talk (e.g., discussing numbers, shapes, and sizes), which enhances children's math skills (Levine et al., 2010). Simple activities like board games and card games also effectively teach young children math concepts (Siegler & Ramani, 2008; Scalise et al., 2018). Caregiver knowledge of math development influences learning outcomes; for instance, educating parents of 4-year-olds about the home numeracy environment increased math-related opportunities at home, improving children's numeracy skills (Niklas et al., 2016). However, parental knowledge is often limited, with many viewing math as less engaging for preschoolers and requiring more formal instruction than language (Cannon & Ginsburg, 2008). Similarly, preschool teachers' knowledge of math development varies, affecting how often they teach math (Ban et al., 2024).

Cognitive development covers the maturation of children's thinking and knowledge acquisition, supporting skills such as attention, flexible thinking, and persistence. Notably, executive functioning (EF) skills—working memory, inhibitory control, and cognitive flexibility—in kindergarten predict reading, math, and science achievement in second grade (Morgan et al., 2019). Similarly, attention-span persistence at age four forecasts math and reading performance at age 21 (McClelland et al., 2013). Caregivers can enhance cognitive skills, particularly EF, through activities like aerobic exercise and mindfulness practices (Diamond & Lee, 2011).

Awareness of healthy sleep habits also supports development, with greater sleep at 12 and 18 months linked to stronger EF at age two (Bernier et al., 2010). Caregiver beliefs about intelligence malleability further influence outcomes: interventions fostering a growth mindset (viewing intelligence as malleable) encourage behaviors like increased pointing gestures with infants (Rowe & Leech, 2019) and selection of challenging preschool activities (Tian et al., 2023). Higher growth mindset levels also correlate with more frequent family learning activities (Justice et al., 2020).

Socioemotional development refers to the children's abilities to form positive relationships, understand others' emotions, and regulate their own emotions and behaviors. Early socioemotional skills significantly shape long-term outcomes: for instance, 72% of children's attachment styles at 12 months persist 20 years later (Waters et al., 2000). Preschool social skills also predict later social competence and academic progress (Frogner et al., 2022; Hammer et al., 2018).

The literature shows that these skills are malleable; training foster parents in responsive interactions increased parental sensitivity, improving children's attachment security at age nine (Zajac et al., 2020) and cortisol regulation in middle school (Garnett et al., 2020). Caregiver knowledge supports socioemotional growth, with greater understanding of evidence-based parenting strategies linked to enhanced parenting competence (e.g., increased affection) and reduced child internalizing behaviors in 2- to 3-year-olds (Winter et al., 2012). Similarly, variability in preschool teachers' knowledge and beliefs about social-emotional learning influences classroom practices (Zinsser et al., 2014).

Screen media use refers to children's engagement with screens, such as televisions and tablets, and its impact on development. This domain was included due to widespread parental misconceptions about the educational benefits of early media use (Lammers et al., 2022; Leung & Suskind, 2020) and evidence of adverse effects from excessive screen time (Zimmerman & Christakis, 2005).

Screen media intersects multiple developmental domains, notably language development, as virtually any content can be displayed. For instance, limiting screen time in the first two years correlates with improved language skills at age two (Supanitayanon et al., 2020). Although prevalent in early childhood (Rideout, 2021), screen use can yield educational benefits for children over two when guided by developmental science (Hirsh-Pasek et al., 2015b). Caregiver knowledge of screen time's effects, both positive and negative, shapes healthy media habits. For example, a parenting intervention increased maternal understanding of television's impact, indirectly reducing children's viewing 3.5 years later (Delisle Nyström et al., 2021).

Dual language learning refers to children's ability to acquire multiple languages in early childhood and its impact on language and other developmental skills. This domain was included due to the increasing number of U.S. children raised with two or more languages and widespread misconceptions among parents and teachers about how young children learn multiple languages (McCabe et al., 2013; Sawyer et al., 2017).

Strong early language skills are critical for developmental outcomes, and supporting the home language in preschool settings enhances math (Partika et al., 2021) and social-emotional skills (Choi & Shen, 2025). Caregiver beliefs shape dual language development; misconceptions that multilingual exposure harms children reduce home language use (Luo et al., 2025). Educating parents about bilingualism's benefits may foster increased language use with children (Baralt et al., 2020).

Brain development refers to the growth of brain structures, functions, and their developmental timing. This domain was included because understanding brain development may inform caregiving practices that enhance children's learning (Zambo, 2008). Like other domains, brain development is malleable; for instance, a family-based intervention increased parent-child conversational turn-taking, leading to positive neurocognitive outcomes in children (Romeo et al., 2021).

Including this domain was exploratory, aiming to examine whether caregiver knowledge of brain development influences interactions and, consequently, child outcomes. For example, awareness of how early linguistic input shapes language areas in infants' brains may encourage parents to engage more with nonverbal infants. Such knowledge could also benefit early childhood educators by guiding effective teaching strategies (Walsh et al., 2024).

Putting the 8 Domains Together

Across the eight domains, content was organized into three primary constructs: normative development, caregiver input, and predicting developmental outcomes. The normative development construct focused on typical patterns of growth within each domain, including key developmental milestones. In the language domain, for instance, it covered the expected progression of children's babbling, first words, two-word combinations, and sentence formation. The caregiver input construct explored how specific adult behaviors influence children's development, either positively or negatively, within each domain. For example, items within the literacy domain focused on the impact of parent-child shared book reading on children's development. The final construct, predicting developmental outcomes, explored how early abilities within a given domain influence later growth—either within the same domain or across others. In the screen media domain, for example, items covered how children's screen usage affects subsequent outcomes, such as language acquisition and socioemotional development.

Several additional domains and topics were considered for inclusion in the SPEAK but were ultimately excluded for various reasons. Notably, atypical development was omitted, as the tool is designed to assess parents' and caregivers' understanding of typical child development.

However, this tool has the potential to enhance parents' awareness of key developmental milestones, which may encourage them to raise concerns during pediatric well-child visits (Gadomski et al., 2018) and seek early evaluation—ultimately contributing to improved child outcomes.

Physical development—including infant and child growth, nutrition, and general health—was excluded, as these topics are routinely covered during pediatric well visits. For instance, Leung & Suskind (2020) found that pediatricians frequently discussed feeding (78.4% at 1 month; 60.3% at 6 months) and infant weight (69.7% at 1 month; 47.2% at 6 months) during these visits. Similarly, another study found that parents of children aged 4 to 35 months reported discussing breastfeeding with pediatricians (85% for visits under 11 minutes long) over the past year. However, discussions about reading were significantly less frequent (49% for visits under 11 minutes long; Halfon et al., 2011)

Safety issues are also frequently discussed during pediatric visits. One study found that 90% of pediatricians advise parents to place infants on their backs for all independent sleep, a practice shown to reduce the risk of Sudden Infant Death Syndrome (SIDS; Koren et al., 2010). Motor development was excluded from the tool, as research suggests parents generally have a strong understanding of this topic (Rikhy et al., 2010), and pediatricians routinely assess it during well-child visits. The American Academy of Pediatrics recommends formal screenings at 9-, 18-, 30-, and 48-month visits (Noritz et al., 2013—recommendation reaffirmed in 2017 and 2022).

In contrast, topics such as learning to talk, infant cognition, and brain development are discussed far less frequently during pediatric well-child visits. Leung and Suskind (2020) report that these subjects come up in fewer than 24% of visits at 1 and 6 months. Additionally, the U.S. Preventive Services Task Force does not recommend routine screening for speech and language delays in children ages 0 to 5 years unless parents express concerns about their child's development (Siu, 2015).

Building the 8 Domains

Separate literature reviews were conducted for each of the eight domains between October 2020 and September 2021. These reviews focused on peer-reviewed journal articles, each summarized in a few sentences by the research team. Studies were included based on the following criteria: (1) participants ranged from 0 through 5 years old, (2) the research examined child development within one of the eight domains of interest, (3) the study provided data on child outcomes rather than solely parental outcomes, and (4) the publication was in English. While no restrictions were placed on the year of publication, the research team ensured that older findings remained supported by current literature.

All citations were collected using Zotero software, and ultimately, over 1,500 articles were reviewed and summarized—more than 650 of which were used as direct sources for item development. Once the literature reviews were complete, the most relevant findings were consolidated into individual PowerPoint presentations for each domain. These presentations were then shared with domain experts to gather feedback on the domain's scope (see Expert Feedback and Content Validity section below).

Item Bank Development

Items for the SPEAK were developed through a rigorous process informed by domain-specific literature reviews and validated measures. Drafted by the primary research team and PhD student assistants, items were refined through multiple reviews by the team and a PhD-level external consultant specializing in child development. Each item was grounded in at least one peer-reviewed journal article, with a subset adapted from the SPEAK (Suskind et al., 2018; Leung & Suskind, 2020) and others informed by measures like the Knowledge of Infant Development Inventory (MacPhee, 1981), Ages and Stages Questionnaire (Squires & Bricker, 2009), and CDC Developmental Milestones (Zubler et al., 2022).

Inclusion criteria required items to focus on child development (birth through 5 years), align with a developmental domain, relate directly to child outcomes, and be consistently supported by literature. To ensure cultural relevance, items were crafted to avoid bias, such as phrasing attachment-related items to account for cross-cultural variations in caregiver roles (Mesman et al., 2016).

Items adhered to best practices in test construction, using clear, concise language (average 16.67 words, SD = 4.34, range: 5–30) and a 7.5 Flesch-Kincaid Grade Level. Written as true/false statements based on research, items used a consistent 4-point Likert scale (definitely not true, probably not true, probably true, definitely true) to capture nuanced responses.

III. Cognitive Interviewing

Cognitive interviewing aimed to ensure SPEAK items were clear and comprehensible to a diverse general population. Participant feedback during interviews highlighted issues, such as unfamiliar or ambiguous terms, guiding item revisions to promote uniform understanding across varied demographics (Willis, 2015). This iterative process enhanced item clarity and accessibility. The study, approved by the University of Chicago’s Biological Sciences Division Institutional Review Board (Protocol: IRB21-0355), obtained verbal consent from all participants. Data collection spanned July 2021 to May 2022.

Participants

Participants were primarily recruited through the University of Chicago's Center for Decision Research (CDR), an online database offering behavioral research studies to a diverse pool of students and non-students. Additional recruitment targeted under-resourced Chicago communities via local organizations, supplemented by snowball sampling and flyers posted locally and on social media. Most data were collected online via Zoom, with a small subset gathered in-person at CDR's Mindworks lab in Chicago. Eligible participants were aged 18+, U.S. residents, and fluent English speakers, verified through a prescreen survey. Compensation was a $20 Amazon gift card per interview, up to $100 for multiple sessions.

The study included 296 adults, with 293 providing demographic data (M_age = 35.74 years, SD = 13.10, range = 18–77). Participants were 57.00% female, 39.25% male, 2.39% other (1.37% no response), and primarily White (45.73%), Black (22.53%), or Asian (21.84%). Most spoke English as their primary language (94.54%) and held a Bachelor's degree or higher (78.16%). Income ranged from under $10,000 (4.78%) to over $200,000 (5.12%). The sample comprised 121 parents (41.30%) and 166 non-parents (56.66%; 2.05% no response), with parents (53.72% mothers, 45.45% fathers) having 1–4 children, mostly one (39.67%) or two (42.15%). Of parents, 47.93% had children aged 0–5 years, 48.76% had children 6+ years (3.31% no response). Detailed demographics are contained in the Appendix.

Repeat Participants

Select participants were invited via email for additional interviews, up to a maximum of five, unless they provided limited responses, struggled with probes, or declined further contact. Re-invitations targeted specific demographics (e.g., educational attainment, gender, parent status) to ensure a diverse sample tested all items comprehensively. In total, 556 interviews were conducted, balancing representation across key demographic groups to validate SPEAK item clarity and cultural relevance.

Procedure

Interviews were scheduled for 1-hour sessions, with durations varying based on the number of SPEAK items presented and participants' response speed to probes. Each interview began with a brief rapport-building conversation, followed by an explanation of the study's purpose. For first-time interviewees, the interviewer read a full consent form, invited questions, and obtained verbal consent (in this manner, this has the spirit of a framed field experiment (List, 2025)). Participants in subsequent interviews received an abbreviated consent script and provided verbal consent. Before the main interview, first-time participants completed a brief, unrelated survey on societal family supports, which is not analyzed here.

Item Review Procedure

After introducing the cognitive interviewing procedure and guiding participants through a warm-up activity, researchers initiated the main interview. Participants were shown items from one or multiple domains of child development, selected based on (1) their demographics—to ensure diverse representation—and (2) their previous exposure to items in earlier interviews. Each item underwent 1 to 5 waves of testing, meaning it was revised and retested based on participant feedback. The minimum number of participants per wave varied: at least 6 for wave 1, 5 for wave 2, and 4 for subsequent waves.

To prevent redundancy, research coordinators ensured that participants in follow-up interviews did not receive items they had previously encountered. On average, participants viewed 18.26 items per interview (SD = 6.80, Range = 3–41). In total, 817 unique items were tested. Items were displayed one at a time, each accompanied by four standardized response options: *definitely not true, probably not true, probably true, definitely true.* Participants read each item aloud, selected their response, and were then prompted to answer general and specific probes about the item before proceeding.

All interviews were audio and video recorded, unless participants requested otherwise. Interviewers took live notes during the sessions, and when real-time note-taking was not feasible, they reviewed recordings afterward to document responses accurately.

Demographic survey

Once participants had completed all assigned items or reached the end of the allotted time, they were instructed to fill out a brief demographic survey. This survey was administered only to those completing their first interview and included questions on basic demographic factors such as age, gender, race, income, education level, and parental status.

Analysis and Revision

Regular meetings were held with the project PI and interviewers to review item feedback from participant interviews. If no interviewer had encountered confusion or issues with an item, the discussion proceeded to the next item. When an issue was identified, the interviewer who observed the concern shared their notes and insights, prompting a group discussion on potential revisions. One team member documented key discussion points to guide the revision process.

Members of the main research team carried out item revisions. Decisions to revise or omit an item were based on participant feedback, insights from similar items, and relevant research findings. Once revisions were finalized for a given wave within a domain, updated items were tested with new participants. If a revision was deemed minor, the item was not re-tested.

IV. Expert Feedback and Content Validity

Expert Selection

The expert panel was composed of three distinct groups of professionals: (1) domain expert researchers, (2) early career researchers who had earned their doctorate in psychology or education within the past 10 years, and (3) Speech-Language Pathologists (SLPs). All experts were invited via email to provide feedback.

Eight domain experts evaluated items within their area of expertise, as well as those overlapping with other domains. Their experience in child development research ranged from 16 to 49 years. Item reviews were conducted through individual 1-hour Zoom meetings, where the research team presented and discussed the items with each expert. Experts provided feedback on various aspects, including the relevance of items to the domain, milestone timing (if applicable), and any concerns regarding content or phrasing.

The early career researcher panel consisted of three researchers who were 1 to 9 years post-PhD in child development research. Each had studied and/or worked across multiple domains within the field and possessed extensive knowledge of recent literature on these topics. Each researcher independently reviewed all items across domains, rating them on a 1-to-4 scale to indicate relevance (1 = not relevant, 2 = slightly relevant, 3 = relevant, 4 = highly relevant). Items rated below 3 required written feedback explaining why they were deemed less relevant.

At the end of each domain review, researchers responded to the following question: *"The items in this survey were intended to measure what is currently known from available research evidence about [domain], with a focus on normative development, effects of caregiver input, and links from early skills to later developmental outcomes. In your opinion, do the items in this survey sufficiently cover this topic? If no, what items should be added to fully capture what is known in* *the research literature about this topic?"* These insights helped the research team identify any gaps and determine additional topics for inclusion.

The SLP panel consisted of five Speech-Language Pathologists at various career stages (4–40 years of experience). Each SLP independently reviewed items within the Language, Literacy, and Dual Language domains, along with any items overlapping with the Language domain. SLPs rated each item on a 1-to-4 scale to assess its relevance to the Language domain (1 = not relevant, 2 = slightly relevant, 3 = relevant, 4 = highly relevant). Items rated below 3 required written feedback explaining the rating. Since SLPs specifically evaluated relevance to the Language domain, only items from this domain were analyzed—excluding those from Literacy, Dual Language, and overlapping categories. Following the independent reviews, the research team held follow-up discussions with three of the SLPs to further examine their feedback.

Lastly, three pediatricians (with 35–41 years of child development expertise) were consulted to verify milestone timing and provide feedback on SPEAK's practical applications. Distinct from the content validity panel, they reviewed items via PowerPoint during a Zoom meeting with the research team, offering insights on developmental milestones and potential clinical use cases to enhance the tool's relevance and accuracy.

Content Validity Results and Expert Feedback

For content validity, we had domain expert researchers review 65–127 items each, rating 98.43%–100% as relevant to their respective child development domains. Experts provided 79 suggestions, leading to 46 item revisions, 10 item omissions, and 18 new items added to the

SPEAK item bank. These changes enhanced the tool's alignment with current research and domain-specific accuracy.

When considering feedback from experts that formally rated the relevance of items (SLPs and early career researchers), content validity was calculated using the index of content validity (CVI) (Lynn, 1986; Polit et al., 2007). CVI is the proportion of experts who gave an item a rating of 3 or 4 for relevance (on a scale from 1- not relevant to 4- highly relevant). To control for possible agreement due to chance, multiple experts (3 to 7 experts) saw each item and a 4-point rating scale (4- highly relevant, 3- relevant, 2- slightly relevant, 1- not relevant) was used, to minimize the likelihood of experts selecting a neutral middle option (Lynn, 1986). CVI was calculated at the item, domain, and full tool levels. At the item level, CVI (I-CVI) ranged from 0.00 to 1.00.

Next, an average CVI was calculated for each domain. Mean CVIs ranged from .93 to 1.00 across domains (Brain: *M* = 0.94, *SD* = 0.13, n = 3; Screen Media: *M* = 0.94, *SD* = 0.20, n = 3; Language: *M* = 0.93, *SD* = 0.12, n = 6-7; Literacy: *M* = 0.98, *SD* = 0.07, n = 3; Socioemotional: *M* = 0.98, *SD* = 0.10, n = 3; Cognitive: *M* = 0.98, *SD* =0.08, n = 3; Dual Language: *M* = 1.00, *SD* = 0.00, n = 3; Math/Spatial: *M* = 1.00, *SD* = 0.00, n = 3). Each of these domain's average CVIs was sufficiently above the .90 threshold for establishing content validity (Polit & Beck, 2006). Across all domains, the average CVI of all items (S-CVI) was .96 (*SD* = .11), also above the .90 threshold.

According to Lynn (1986) an additional method to control against chance agreement is to ensure CVI is beyond the .05 level of significance. Polit et al. (2007) similarly suggest that CVI values over .78 are considered valid. Using both Lynn's (1986) conventions and Polit et al.'s (2007) cut-off values (both identified the same items), there were 7 items in the Brain domain, 7 in the Screen Media domain, 13 in the Language domain, 4 in the Literacy domain, 2 in the Socioemotional domain, and 4 in the Cognitive domain judged to be not content valid. There were no items judged to be not content valid in the Dual Language or Math/Spatial domains. Overall, 523 of 560 items (93.39%) were judged content valid using this more stringent criteria.

As a result of this review, 3 items were revised and 4 items were omitted. The remaining items were retained for a variety of reasons, including the feedback from domain expert researchers and comments from the raters that indicated the item was not relevant to the specific domain but was relevant to overall knowledge of child development.

Lastly, two pediatricians reviewed SPEAK items, with one evaluating 25 and the other 56. They provided suggestions for 11 items, leading to 4 revisions and 4 omissions. The remaining 3 items were retained without changes, informed by other expert advice and literature review. No new topics were proposed.

V. Calibration Study

This section outlines our calibration study undertaken to inform item selection and parameter estimation for the SPEAK using item response theory (IRT).

Participant Sample

Participants were recruited through Prolific, an online platform that connects volunteer participants with research studies. A total of 1,100 participants contributed data to the calibration study, with 1,013 providing usable responses. Data exclusions were made for various reasons, including incomplete responses, failure of attention checks, unrealistically short completion times (under 9 minutes for smaller item subsets; under 12 minutes for half-item sessions), suspected bot activity, demographic discrepancies between Prolific records and study-specific surveys, and multiple submissions for the same items. In total, 87 participants (7.9% of the sample) were excluded.

Among usable participants, the average age was 42.09 years (SD = 13.95, range: 18–83), with 51% identifying as female. The sample was 72.1% White, and 84.2% identified as not Hispanic or Latino. Regarding education, 65.2% had obtained less than a Bachelor's degree. Parents made up 66.9% of the sample, and 33.1% had a child 5 years old or younger.

For a more detailed breakdown of demographic characteristics, see Table 1. Data collection took place from February 2023 to May 2023, and the University of Chicago Social Sciences Institutional Review Board (IRB22-1117) approved the study with a waiver of written informed consent. Participants reviewed an online consent information sheet and indicated their willingness to participate by selecting a checkbox to continue.

Table 1. Calibration Study Participant Demographics

Demographic Characteristic	N	Percent of Sample
Gender		
Male	473	46.7%
Female	517	51.0%
Transgender Male	2	.2%
Transgender Female	6	.6%
Another gender	13	1.3%
Prefer not to answer	2	.2%
Ethnicity		
Not Hispanic/Latino	853	84.2%
Hispanic/Latino	156	15.4%
Prefer not to answer	4	0.4%
Race		
White	730	72.1%
Black/African American	142	14.0%
Asian	42	4.1%
American Indian/Alaskan Native	12	1.2%
Native Hawaiian/Other Pacific Islander	1	.1%
Multiracial	52	5.1%
Other	23	2.3%
Prefer not to answer/no response	11	1.1%

Education		
8th Grade or less	2	.2%
9th-12th Grade, no diploma	31	3.1%
High school graduate	141	13.9%
GED	47	4.6%
Postsecondary non-degree program	24	2.4%
Some college credit but not degree	288	28.4%
Associates Degree (AA, AS)	128	12.6%
Bachelor’s Degree (BA, BS)	214	21.1%
Some graduate school credit but no degree	17	1.7%
Master’s Degree (MS, MBA)	98	9.7%
Doctorate or Professional Degree	23	2.3%
Income		
Under $10,000: 52 (5.1%)	52	5.1%
$10,000-19,999: 80 (7.9%)	80	7.9%
$20,000-29,999: 117 (11.5%)	117	11.5%
$30,000-39,999: 112 (11.1%)	112	11.1%
$40,000-49,999: 116 (11.5%)	116	11.5%
$50,000-74,999: 219 (21.6%)	219	21.6%
$75,000-99,999: 110 (10.9%)	110	10.9%
$100,000-149,999: 124 (12.2%)	124	12.2%
$150,000-199,999: 49 (4.8%)	49	4.8%
Over $200,000: 34 (3.4%)	34	3.4%
Number of children		
0 children (non-parent)	335	33.1%
1 child	227	22.4%
2 children	250	24.7%
3 children	117	11.5%
4 children	59	5.8%
5 children	11	1.1%
6 children	10	1.0%
7 children	3	.3%
8 children	1	.1%
Age of youngest child		
Under 1 year	45	4.4%
1 year	57	5.6%
2 years	68	6.7%
3 years	74	7.3%
4 years	54	5.3%
5 years	39	3.8%
6 years	26	2.6%
7 years	15	1.5%
8 years	16	1.6%

9 years	11	1.1%
10 years	14	1.4%
11 years	16	1.6%
12 years	19	1.9%
13 years	16	1.6%
14 years	11	1.1%
15 years	9	0.9%
16 years	12	1.2%
17 years	10	1.0%
18 years or older	159	15.7%
No response/No children	342	33.8%

Item Presentation

The calibration study included usable data from 1,013 participants. Of these, 302 completed the full SPEAK item bank (560 items) over two sessions (1–20 days apart), while 711 completed a subset via a balanced incomplete block design (BIBD) in a single session.

In the BIBD, items were divided into 28 blocks of 20 items, balanced by domain, construct (normative, caregiver input, predicting outcomes), age group, and word count. These blocks were organized into seven forms, each containing four blocks (80 items per form). Participants completed two of the seven forms (160 items), plus 15 SPEAK-derived items and nine experimental items with a dichotomous scale. Form pairings (22 combinations) and presentation order were counterbalanced.

Participants were stratified by education (aligned with U.S. Census: 9.4% lowest, 54.7% middle, 35.9% highest) and parent status (parents of 0–5-year-olds, 6+-year-olds, or non-parents), then randomly assigned to one of 22 BIBD groups or the full item bank group (in the parlance of experimental economics this is denoted as ‘blocking’ or ‘stratification’).

Attention Checks

Attention checks were implemented to ensure data quality. Participants who completed the full item bank encountered six attention checks (three per session), consisting of two nonsensical items that presented impossible scenarios (e.g., *"Most newborn infants (0 to 1 month) have been to every country in the world."*) and one instructional manipulation check (e.g., *"Please select 'Probably True' for this question. This is an attention check."*). Participants tested on a subset of items saw two nonsensical attention checks, where acceptable responses were 'definitely not true' or 'probably not true.' For these participants, correct responses on all attention checks were required for their data to be considered usable. Given the longer testing duration for participants who completed the full item bank, data was deemed usable if at least four of six attention checks were answered correctly—provided no other concerns were present.

Empirical Results

Calibration

Results from the calibration study indicated that a unidimensional graded response model (GRM) provided the best fit for the data. The model was theoretically appropriate, as higher parental knowledge was expected to correlate with higher scores on the scale. Confirmatory factor analysis was conducted to assess whether items loaded onto the original eight domains (e.g., language development, math and spatial development), but the model failed to converge. Exploratory factor analyses using two- and three-factor models were also tested; however, item groupings lacked strong theoretical rationale. Ultimately, the data best supported a unidimensional structure, with all items loading onto a general parenting knowledge factor—consistent with findings from the original SPEAK measure (Suskind et al., 2018).

Item retention decisions were based on sufficient discrimination (minimum = .35) and independence (no local dependence). When two items exhibited dependence, the item with higher discrimination was retained. If dependent items had similar discrimination values, researchers prioritized the item that was theoretically more relevant to the main construct. In total, 404 items were retained in the model, with an average item discrimination of .918 (SD = .363; range: .350–2.45).

Simulated CAT

A simulated Computerized Adaptive Test (CAT) was used to evaluate SPEAK's average test length, ensuring content balance by including at least one item from each of eight domains and a minimum of eight items overall. A 30-item maximum maintained a user-friendly experience. The initial item was randomly selected from the full item bank to maximize exposure (i.e, a randomesque approach), with subsequent items chosen randomly from the five most appropriate items using maximum Fisher information (MFI) and Bayes Modal (BM) for ability estimation. For a standard error (SE) of 0.3 (reliability ≈ 0.91; Fenwick et al., 2023), the average test length was 23.11 items (range: 8–30). Age-group-specific CAT simulations, using identical criteria, required 27.35–29.23 items on average (range: 20–30).

VI. Validation Study

This section outlines our approach to validating the SPEAK, with a focus on examining concurrent links with parent behaviors and child outcomes.

Participant Sample

Participants were primarily recruited through three online participant platforms: ResearchMatch (N = 137), Prolific (N = 80), and University of Chicago's CDR (N = 46). Participants were also recruited through emails sent to parent-specific listservs (N = 14), social media (N = 12), and flyers posted at the University of Chicago (N = 2). Eligibility criteria required participants to be at least 18 years old, live in the U.S., be fluent in English, and have at least one child between 0 and 5 years old. These criteria were confirmed using a brief eligibility screener administered prior to enrollment. In total, 291 participants contributed useable data, and 91 participants (23.8% of the total sample) were excluded for failure of attention checks (N = 58), incomplete responses (N = 25), and suspected duplicate submissions (N = 8).

Among usable participants, the average age was 33.85 years (SD = 6.05, range: 20–61), with 82.5% identifying as female. The sample was 61.9% White, and 82.1% identified as not Hispanic or Latino. Regarding education, 52.6% had obtained less than a Bachelor's degree. All participants had at least one child 5 years old or younger; 32.0% of participants had two or more children in this age range.

For a more detailed breakdown of demographic characteristics, see Table 2. Data collection took place from March 2024 to May 2025, and the University of Chicago Social Sciences Institutional Review Board (IRB23-1743) approved the study with a waiver of written informed consent. Participants reviewed an online consent information sheet and indicated their willingness to participate by selecting a checkbox to continue.

Table 2. Validation Study Participant Demographics

Demographic Characteristic	N	Percent of Sample
Gender		
Male	50	17.2%
Female	240	82.5%
Another gender	1	0.3%
Ethnicity		
Not Hispanic/Latino	239	82.1%
Hispanic/Latino	51	17.5%
Prefer not to answer	1	0.3%
Race		
White	158	54.3%
Black/African American	82	28.2%
Asian	11	3.8%
American Indian/Alaskan Native	2	.7%
Native Hawaiian/Other Pacific Islander	2	.7%
Multiracial	26	8.9%
Other	7	2.4%

Prefer not to answer/no response	3	1.0%
Education		
8th Grade or less	1	0.3%
9th-12th Grade, no diploma	1	0.3%
High school graduate	36	12.4%
GED	12	4.1%
Postsecondary non-degree program	9	3.1%
Some college credit but not degree	56	19.2%
Associates Degree (AA, AS)	38	13.1%
Bachelor's Degree (BA, BS)	63	21.6%
Some graduate school credit but no degree	8	2.7%
Master's Degree (MS, MBA)	50	17.2%
Doctorate or Professional Degree	17	5.8%
Income		
Under $10,000: 52 (5.1%)	8	2.7%
$10,000-19,999: 80 (7.9%)	16	5.5%
$20,000-29,999: 117 (11.5%)	32	11.0%
$30,000-39,999: 112 (11.1%)	32	11.0%
$40,000-49,999: 116 (11.5%)	21	7.2%
$50,000-74,999: 219 (21.6%)	61	21.0%
$75,000-99,999: 110 (10.9%)	47	16.2%
$100,000-149,999: 124 (12.2%)	42	14.4%
$150,000-199,999: 49 (4.8%)	18	6.2%
Over $200,000: 34 (3.4%)	10	3.4%
Prefer not to answer	4	1.4%
Number of children aged 0-5		
1 child	198	68.0%
2 children	84	28.9%
3 children	7	2.4%
4 children	2	0.7%
Age of youngest child		
Under 1 year	67	14.8%
1 year	44	9.7%
2 years	54	11.9%
3 years	51	11.3%
4 years	46	10.2%
5 years	29	6.4%

Validation Measures

In addition to administering the SPEAK to assess parental knowledge of child development, several other validated or widely used measures of parent behavior and child outcomes were administered to support analyses of concurrent validity. These instruments included a combination of parent interviews, parent surveys, and direct child assessments.

KIDI

The Knowledge of Development Inventory (KIDI) is a survey that has been used to evaluate parenting knowledge, focusing on developmental processes, developmental milestones, parenting practices, and health and safety (MacPhee, 1981). A 10-item simplified version of the KIDI was used, in which parents were asked whether they agreed, disagree, or were not sure about statements regarding child development (e.g., "All infants need the same amount of sleep"; Rowe et al., 2016). Six of the items focused on milestones (e.g., "A one-year-old knows right from wrong"), and if a parent disagreed with one of these statements, a follow up question asked whether the parent thought children have the ability at a younger or older age. Prior research with this 10-item version showed evidence of validity and acceptable reliability (α=.60) comparable to other short versions of the KIDI (Rowe et al., 2016), however, reliability was lower in the current sample (α=.55).

HFPI

The Healthy Families Parenting Index (HFPI) is a reliable and valid 63-item survey tool that includes nine subscales covering different aspects of parenting practices and family health/functioning (Krysik & Lecroy, 2012). Items were rated on a 5-point Likert scale from "Rarely or Never" to "Always or most of the time," and items in each scale were summed to create subscale scores. All subscales demonstrated strong internal reliability (α=0.82 to 0.93). Two of the subscales, *Parent/Child Interaction* (e.g., "I respond quickly to my child's needs.") and *Home*

Environment (e.g., “I have organized my home for raising a child.”) focus on parenting behaviors at the family level and were expected to show evidence of convergent validity with the SPEAK.

Subscales focused on the parent as an individual were expected to show evidence of divergent validity with the SPEAK. These included subscales on *Personal Care* (e.g., “I get enough sleep.”), *Depression* (e.g., “I feel sad.”), *Parenting Efficacy* (e.g., “I am proud of myself as a parent.”), *Role Satisfaction* (e.g., “Because I’m a parent, I’ve had to give up much of my life.”), and *Problem Solving* (e.g., “I am good at dealing with unexpected problems.”). Similarly, subscales at the community/societal level, specifically *Social Support* (e.g., “I discuss my feelings with someone.”) and *Mobilizing Resources* (e.g., “I know where to find resources for my family.”) were expected to show evidence of divergent validity with the SPEAK.

STIMQ$_2$

The STIMQ$_2$ parent interview assessed the cognitive environment provided by the parent, specifically the caregiver’s fostering of learning and development through book reading, teaching activities, and verbal responsivity (Cates et al., 2023). Three developmental versions (Infant, Toddler, and Preschool) were used depending on the age of the target child. Across the three versions, three subscales were included: Reading (READ; books /reading activities), Parental Involvement in Developmental Advance (PIDA; teaching activities) and Parental Verbal Responsivity (PVR; verbal interactions). The fourth subscale, the Availability of Learning Materials (ALM; variety of toys), was not administered as it is not included in the core cognitive stimulation score. The three age versions of the STIMQ$_2$ demonstrate moderate to strong internal consistency and validity with measures of child language, social-emotional skills, and cognitive ability (Cates et al., 2023). In the current sample, internal consistency was strong for all three age versions (Infant: α=0.90, Toddler: 0.85, and Preschool: 0.93).

PROMIS-Early Childhood

The Patient-Reported Outcome Measurement Information System Early Childhood (PROMIS EC) is a concise survey tool that evaluates various aspects of young children's well-being (Cella et al., 2022). The tool has been normed to the U.S. population, providing standardized T-scores. Four PROMIS EC scales were used to capture children's social-emotional and cognitive development: 1) *Self-regulation: Flexibility,* capturing children's ability to adapt to changes (5 items, α=.84), 2) *Social-relationships: Child-caregiver Interactions*, capturing children's positive social behaviors and connectedness with their caregiver (5 items, α=.82), 3) *Engagement: Persistence*, capturing children's ability to sustain effort in problem solving and complete challenging activities (6 items, α=.84), and *Engagement: Curiosity*, capturing children's initiative and interest in exploring new things (6 items, α=.70). These measures all used 5-point Likert scales from "Never" to "Always" and asked parents to respond about the last 7 days.

ASQ:SE

The Ages and Stages Questionnaire: Social Emotional (ASQ:SE) was used as a broader measure of social-emotional behaviors across age stages (Squires et al., 2002). Parents responded to a series of age-specific questions such as "Does your baby smile at you and other family members?" on a three-point scale from "Never" to "Most of the time." For each question, parents could also indicate whether the behavior was a concern for them. The ASQ:SE is used as a screening tool by pediatricians to identify young children who may require further assessment or intervention for social or emotional delays, when scores fall below a certain cutoff. Analyses of the ASQ:SE data are not included in this paper due to insufficient power to detect effects, given the small sample sizes for each age form (range: 4-25). Additional data on the SPEAK and ASQ:SE are being collected as part of a different study and will be analyzed separately.

CDI-CAT

The language skills of children ages 15 to 36 months were assessed using the MacArthur Bates Communicative Development Index—Computer Adaptive Test (CDI-CAT), a parent report measure of children's expressive vocabulary (Kachergis et al., 2022). Parents responded as to whether their child understands and says a personalized list of words, with each word selected adaptively based on their responses to previous items. The CDI-CAT was developed using data from long-form CDIs (i.e., comprehensive measures of language development with demonstrated validity and reliability); a validation study supported a strong correlation ($r = 0.92$) between the CDI-CAT and the CDI: Words and Sentences (Dale, 1991; Kachergis et al., 2022).

WJ IV Picture Vocabulary

The language skills of children ages three through five years were measured using the Picture Vocabulary subtest of the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank & Wendling, 2018). This assessment required children to identify pictured objects, using pointing for the first two items and by verbally naming the remaining items. The test was ended when children answered six consecutive items incorrectly.

Verbal Counting Task

Counting ability was assessed by asking children to count aloud as high as possible, up to a maximum of 50. The largest number that children counted to correctly (i.e., without a sequencing error) was recorded. Research suggests verbal counting skills, assessed in this way in early childhood, predict school math performance years later (Hannula-Sormunen et al., 2015).

How-Many Cardinality Task

Children's cardinality knowledge was measured using the How-Many task (adapted from Wynn, 1992 and O'Rear et al., 2024), in which children were shown stars linearly arranged on the screen and asked to count them. After the child finished counting, the stars were hidden and children were asked how many stars there were (i.e., how many were hiding). Children were first shown a practice trial with 2 stars, and were then shown test sets of 3, 4, 8, 16, and 20 stars presented in order of increasing magnitude. The percentage of trials in which the child both counted correctly and correctly stated the number of hidden stars was recorded (internal consistency: α=.83).

Procedure

Caregivers were invited to participate in up to three sessions. The first session, a prerequisite for either of the other sessions, involved the completion of a set of surveys, including the SPEAK, KIDI, HFPI, and a demographic questionnaire. Two attention checks were included as part of this session. If attention checks were passed and if the parent had a child 5 months or older, an invitation was made for the next session.

The second session, conducted live with a researcher over Zoom, included a 10-minute free play session involving the parent and child (not analyzed here) and, for children ages three to five years, the WJ IV Picture Vocabulary test, verbal counting task, and the how-many cardinality task. Caregivers also completed surveys following the live session, which varied depending on the age of the child. Caregivers of children aged 5 to 14 months completed the ASQ:SE; caregivers of children aged 15 to 36 months completed the ASQ:SE, CDI-CAT, and PROMIS EC scales; caregivers of children aged three to five years completed the ASQ:SE and PROMIS EC scales.

The third session was the $STIMQ_2$ parent interview, conducted with a member of the research team over Zoom. Caregivers recruited from Prolific were only invited to the first and third sessions; they were not invited to the parent-child session because Prolific policy prohibits minors from participating in their studies.

Empirical Results

Administration Times

The median time required to complete the SPEAK was 4.35 minutes (interquartile range [IQR] = 3.18-6.33). The median number of items completed was 24 (range: 17-30).

Demographic Predictors

Analyses of variance (ANOVA) were used to test for significant group differences across three demographic variables that were predicted to be linked with knowledge of child development: parent gender, education level, and coursework/training specific to early childhood. There was a significant effect of gender on SPEAK scores, $F(1,288) = 6.69$, $p = .01$, Cohen's $d = .40$, with mothers demonstrating greater knowledge of child development ($M = 0.44$, $SD = 0.86$) than fathers ($M = 0.09$, $SD = 0.92$). No effect of education level on SPEAK scores was found, $F(4,286) = 1.54$, $p = .19$, possibly reflecting the limited variability in education within this relatively highly educated sample. However, there was a significant effect of coursework/training specific to early childhood, $F(1,289) = 13.73$, $p < .001$, Cohen's $d = .54$. Parents with formal experience in this field demonstrated greater knowledge ($M = 0.74$, $SD = 0.72$) than parents with no such experience ($M = 0.28$, $SD = 0.89$).

Concurrent Validity

Pearson's correlations between SPEAK scores and other measures of concurrent validity are presented in Table 3. The pattern of correlations provides evidence of convergent and divergent validity. Specifically, as expected, SPEAK scores were significantly and positively associated with parental knowledge and behavior as measured by the KIDI, HFPI Parent/Child Interaction, HFPI Home Environment, and $STIMQ_2$.

Also consistent with predictions, SPEAK scores did not significantly correlate with HFPI subscales on Personal Care, Depression, Parental Efficacy, Role Satisfaction, or Social Support. Contrary to hypotheses, however, SPEAK scores showed significant associations with the HFPI Problem Solving and Mobilizing Resources subscales. These unexpected correlations may suggest that parents' resourcefulness is linked to the extent to which they seek or acquire knowledge, though future research is needed to clarify the direction of this association.

SPEAK scores were also significantly associated with several child outcome measures, including the WJ IV Picture Vocabulary and the PROMIS EC scales of Flexibility, Child-caregiver Interactions, Persistence, and Curiosity. Associations with the Verbal Counting Task and CDI-CAT approached significance, whereas no significant relation emerged between SPEAK scores and the How-Many Cardinality Task.

Table 3. Concurrent relationships between SPEAK scores and measures of parent knowledge, parent behaviors, and child outcomes

Measure	**N**	***r***	***p***
KIDI	291	.24	<.001
HFPI Parent/Child Interaction	291	.23	<.001
HFPI Home Environment	291	.30	<.001
HFPI Personal Care	291	.06	.33
HFPI Depression	291	.10	.10
HFPI Parent Efficacy	291	.09	.12
HFPI Role Satisfaction	291	-.03	.61
HFPI Problem Solving	291	.27	<.001
HFPI Social Support	291	.09	.11
HFPI Mobilizing Resources	291	.21	<.001
$STIMQ_2$	120	.34	<.001

PROMIS EC Flexibility	94	.29	.005
PROMIS EC Child-caregiver Interactions	94	.38	<.001
PROMIS EC Persistence	94	.36	<.001
PROMIS EC Curiosity	94	.25	.02
WJ IV Picture Vocabulary	57	.35	.008
How-Many Cardinality Task	56	.16	.24
Verbal Counting Task	55	.23	.09
CDI-CAT	31	.32	.07

Note. KIDI = Knowledge of Development Inventory (α=.55); HFPI = Healthy Families Parenting Index; PROMIS EC = Patient-Reported Outcome Measurement Information System Early Childhood; CDI-CAT = Computerized Adaptive Test version of the Communicative Development Inventories; WJ = Woodcock-Johnson. Internal reliability (Cronbach's α) was .55 for KIDI and ranged from .70-.93 for other scales.

Exploratory analyses examined whether raw scores from SPEAK domains showed domain-specific associations with child math outcomes. Results suggested that scores on the math domain, but not other domains, were significantly associated with the Verbal Counting Task and the How-Many Cardinality Task (Table 4). These findings are preliminary and should be interpreted with caution, as the SPEAK was designed as a unidimensional measure and domain-level analyses were not planned.

Table 4. Correlations between SPEAK domain raw scores and child math outcomes

Child Math Outcome	Language Domain	Literacy Domain	Math Domain	Socio-emotional Domain	Cognitive Domain	Brain Domain	Dual Language Domain	Screen Media Domain
How-Many Cardinality Task	.16	.10	.34*	.11	.12	.18	-.13	.07
Verbal Counting Task	.21	.20	.29*	.21	.10	.11	.03	.07

Note. *$p < .05$. Analyses were exploratory and domain scores were calculated as the average of raw item scores within each domain.

VII. Future Directions

The SPEAK, a fully operational computer-adaptive tool, is already in use by research and non-profit organizations across the U.S. and internationally, including active collaborations and data collection in Saudi Arabia and Thailand. Its uptake across diverse settings underscores its utility, and preliminary evidence of regional and socioeconomic differences highlights its potential for generating new insights. These efforts are accompanied by ongoing validation work, ensuring the tool's applicability across diverse languages and cultural contexts.

In addition to providing overall SPEAK scores, some information about results at the domain-level may also be beneficial to users and stakeholders. Though the test is unidimensional, exploratory analyses of the validity data suggested some domain-specific associations. Further, more qualitative information about performance in domains could help inform targeted education, such as professional development opportunities for early childhood educators. Given that providing feedback for the eight separate domains may not be useful for actionable next steps (e.g., a low score in the brain domain would likely involve education about practices involving other domains, such as increasing language input), the domains will be condensed into five more digestible topics.

In this more digestible form, the SPEAK assesses five domains: 1) language and literacy, 2) socioemotional development, 3) STEM learning, 4) child-technology interaction, and 5) dual language learning. The new STEM domain emphasizes early math, spatial, and scientific knowledge. The new child-technology interaction domain will be an expansion of the screen media domain, including emerging technologies like artificial intelligence, which can affect learning across domains. Items were reorganized: original language and literacy items were merged, cognitive domain items were reassigned to STEM (math/science) or socioemotional (attention/memory), and brain domain items were redistributed based on focus (e.g., social cognition to socioemotional, language areas to language/literacy). Content balancing ensures at least two items per domain are administered, optimizing coverage across the revised domains.

A key strength of the SPEAK's computer-adaptive design is its flexibility to incorporate new items as child development research evolves, particularly in the emerging area of artificial intelligence (AI). AI is increasingly integrated into early childhood settings, from social robots like QRIO, which toddlers treated as peers in childcare classrooms (Tanaka et al., 2007), to AI chatbots teaching parents skills for managing child behaviors, with 74% skill retention (Entenberg et al., 2021). Yet, the cognitive impacts of early AI exposure remain understudied (Suskind, 2025). New AI-focused items will assess caregivers' knowledge, informing how parents and educators navigate AI's benefits and risks in fostering child development.

VIII. Epilogue

Early education, spanning a child's first five years, is pivotal for lifelong cognitive, social, emotional development, and sets the stage for future economic success. Its transformative value lies in shaping developmental pathways, narrowing inequalities, and yielding lasting societal benefits. However, effective delivery remains a critical challenge, as the quality and consistency of instructional content significantly influence outcomes, necessitating targeted strategies to ensure impactful, scalable programs. We view SPEAK as an exciting innovation that enables parents, teachers, clinicians, policymakers, and researchers to assess child development knowledge efficiently and with precision.

Policymakers may find SPEAK valuable for several key reasons, including:

- Informing Targeted Policy Interventions and Combating Inequities: Policymakers prioritize evidence-based solutions to address societal challenges, and the SPEAK provides a scalable tool to identify gaps in parental and educator knowledge of child development. By pinpointing specific areas of need, particularly in underserved populations with varying educational attainment, it enables the design of targeted policies, such as subsidized training programs or public health campaigns, to enhance caregiving practices and improve child outcomes.
- Cost-Effective Resource Allocation: The SPEAK's efficiency in assessing knowledge quickly and precisely allows policymakers to allocate resources effectively. For example, it can guide investments in clinical or educational interventions where knowledge deficits are most pronounced, maximizing the impact of limited budgets on child development outcomes like language and socioemotional skills, which have long-term economic benefits. The lack of a comprehensive assessment tool has been a major barrier in child development policy.
- Supporting Early Childhood Development Goals: Policymakers are increasingly focused on early childhood development as a foundation for societal well-being. The SPEAK's

ability to measure knowledge linked to better parent-child interactions and child outcomes aligns with policy goals to boost school readiness and reduce developmental disparities, offering a data-driven approach to track progress and evaluate intervention impacts.

For academics, SPEAK might also prove valuable for several reasons, including:

- Advancing Measurement in Developmental Research: Academics in psychology, education, and developmental science will value the SPEAK as a novel, computer-adaptive tool that fills a critical gap in assessing child development knowledge. Its precision and scalability enable researchers to study knowledge variations across populations and over time, facilitating longitudinal studies and intervention evaluations with robust, reliable data.
- Interdisciplinary Relevance and Scaling: The paper bridges multiple disciplines, including developmental psychology, education, and policy research, by linking parental knowledge to child outcomes and offering a practical tool for intervention. Academics will be interested in exploring how SPEAK can be applied in diverse contexts, such as clinical trials, teacher training studies, or socioeconomic analyses, fostering cross-disciplinary collaborations. Importantly, traditional assessments often rely on small-scale studies, but SPEAK's computer-adaptive design enables large-scale investigations across diverse, representative populations. This scalability not only enhances external validity (List, 2022; 2024), but also facilitates cross-national collaborations, supports multi-institutional partnerships, and empowers governments, NGOs, and academics worldwide to generate comparable data and advance early childhood research on a global scale.
- Building on and Extending Existing Literature: The SPEAK builds on validated tools like SPEAK and integrates insights from extensive research (e.g., Hirsh-Pasek et al., 2015a;

List et al., 2021). Academics will appreciate its rigorous methodology and potential to extend the literature by providing a standardized metric to test hypotheses about knowledge, caregiving behaviors, and child development outcomes, encouraging further empirical and theoretical advancements.

In sum, we view the SPEAK as deepening the tool kit of how we assess child development knowledge, offering a precise, scalable, computer-adaptive tool that empowers clinicians, educators, researchers, and policymakers. By swiftly identifying knowledge gaps, it enables tailored interventions, enhances professional development, and informs evidence-based policies, ultimately fostering better parent-child interactions and boosting children's language, cognitive, and socioemotional outcomes. Its rigorous design and broad applicability make SPEAK an indispensable asset for advancing early childhood development and closing disparities in caregiving knowledge. The value of this tool, however, will only arrive upon its broad adoption: please let us know how we can help!

References

Al-Ubaydli, O., List, J. A. & Suskind, D. (2017). Klein Lecture: The Science of Using Science: Toward an Understanding of the Threats to Scalability. *International Economic Review* 61, 1387–1409 (2020). *International Economic Review* 61, 1387– 1409 (2020).

Baralt, M., Darcy Mahoney, A., & Brito, N. (2020). Háblame Bebé: A phone application intervention to support Hispanic children's early language environments and bilingualism. *Child Language Teaching and Therapy*, *36*(1), 33-57.

Bennett, K. K., Weigel, D. J., & Martin, S. S. (2002). Children's acquisition of early literacy skills: Examining family contributions. *Early Childhood Research Quarterly*, *17*(3), 295-317.

Bernier, A., Carlson, S. M., Bordeleau, S., & Carrier, J. (2010). Relations between physiological and cognitive regulatory systems: infant sleep regulation and subsequent executive functioning. *Child development*, *81*(6), 1739-1752.

Blake, J., Macdonald, S., Bayrami, L., Agosta, V., & Milian, A. (2006). Book reading styles in dual-parent and single-mother families. *British journal of educational psychology*, *76*(3), 501-515.

Cannon, J., & Ginsburg, H. P. (2008). "Doing the math": Maternal beliefs about early mathematics versus language learning. *Early education and development*, *19*(2), 238-260.

Cates, C. B., Roby, E., Canfield, C. F., Johnson, M., Raak, C., Weisleder, A., ... & Mendelsohn, A. L. (2023). Validation of the StimQ$_2$: A parent-report measure of cognitive stimulation in the home. *PLoS One*, *18*(7), e0286708.

Cella, D., Blackwell, C. K., & Wakschlag, L. S. (2022). Bringing PROMIS to early childhood: Introduction and qualitative methods for the development of early childhood parent report instruments. *Journal of Pediatric Psychology*, *47*(5), 500-509.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1-29.

Connor, D. O., Kirkland, P. K., & Purpura, D. J. (2024). The how many and give-N tasks: Conceptually distinct measures of the cardinality principle. *Early Childhood Research Quarterly*, *66*, 61-74.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. Journal of Speech and Hearing Research, 34, 565–571.

Delisle Nyström, C., Abbott, G., Cameron, A. J., Campbell, K. J., Löf, M., Salmon, J., & Hesketh, K. D. (2021). Maternal knowledge explains screen time differences 2 and 3.5 years post-intervention in INFANT. *European Journal of Pediatrics*, *180*(11), 3391-3398.

Demir-Lira, Ö. E., Applebaum, L. R., Goldin-Meadow, S., & Levine, S. C. (2019). Parents' early book reading to children: Relation to children's later language and literacy outcomes controlling for other parent language input. *Developmental science*, *22*(3), e12764.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959-964.

Dunst, C. J., Simkus, A., & Hamby, D. W. (2012). Relationship between age of onset and frequency of reading and infants' and toddlers' early language and literacy development. *Center for Early Literacy Learning*, *5*(3), 1-10.

Entenberg, G. A., Areas, M., Roussos, A. J., Maglio, A. L., Thrall, J., Escoredo, M., & Bunge, E. L. (2021). Using an artificial intelligence based chatbot to provide parent training: results from a feasibility study. *Social Sciences*, *10*(11), 426.

Frogner, L., Hellfeldt, K., Ångström, A. K., Andershed, A. K., Källström, Å., Fanti, K. A., & Andershed, H. (2022). Stability and change in early social skills development in relation to early school performance: A longitudinal study of a Swedish cohort. *Early Education and Development*, *33*(1), 17-37.

Gadomski, A. M., Riley, M. R., Scribani, M., & Tallman, N. (2018). Impact of "Learn the signs. Act early." Materials on parental engagement and doctor interaction regarding child development. *Journal of Developmental & Behavioral Pediatrics*, *39*(9), 693-700.

Garnett, M., Bernard, K., Hoye, J., Zajac, L., & Dozier, M. (2020). Parental sensitivity mediates the sustained effect of Attachment and Biobehavioral Catch-up on cortisol in middle childhood: A randomized clinical trial. *Psychoneuroendocrinology*, *121*, 104809.

Golinkoff, R. M., Can, D. 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.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child development*, *90*(3), 985-992.

Halfon, N., Stevens, G. D., Larson, K., & Olson, L. M. (2011). Duration of a well-child visit: association with content, family-centeredness, and satisfaction. *Pediatrics*, *128*(4), 657-664.

Hentges, R. F., Devereux, C., Graham, S. A., & Madigan, S. (2021). Child language difficulties and internalizing and externalizing symptoms: A meta-analysis. *Child Development*, *92*(4), e691-e715.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., ... & Suma, K. (2015a). The contribution of early communication quality to low-income children’s language success. *Psychological science*, *26*(7), 1071-1083.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015b). Putting education in “educational” apps: Lessons from the science of learning. *Psychological science in the public interest*, *16*(1), 3-34.

Hammer, D., Melhuish, E., & Howard, S. J. (2018). Antecedents and consequences of social–emotional development: A longitudinal study of academic achievement. *Archives of Scientific Psychology*, *6*(1), 105-116.

Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool children’s spontaneous focusing on numerosity, subitizing, and counting skills as predictors of their mathematical performance seven years later at school. *Mathematical Thinking and Learning*, *17*(2-3), 155-177.

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American journal of speech-language pathology*, *19*(1), 51-65.

Justice, L. M., Purtell, K. M., Bleses, D., & Cho, S. (2020). Parents’ growth mindsets and home-learning activities: A cross-cultural comparison of Danish and US parents. *Frontiers in psychology*, *11*, 1365.

Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., & Frank, M. C. (2022). Online computerized adaptive tests of children's vocabulary development in English and Mexican Spanish. *Journal of Speech, Language, and Hearing Research*, *65*(6), 2288-2308.

Kim, S., Chang, M., & Kim, H. (2011). Does teacher educational training help the early math skills of English language learners in Head Start?. *Children and youth services review*, *33*(5), 732-740.

Koren, A., Reece, S. M., Kahn-D'angelo, L., & Medeiros, D. (2010). Parental information and behaviors and provider practices related to tummy time and back to sleep. *Journal of Pediatric Health Care*, *24*(4), 222-230.

Krysik, J., & Lecroy, C. W. (2012). Development and initial validation of an outcome measure for home visitation: The healthy families parenting inventory. *Infant Mental Health Journal*, *33*(5), 496-505.

Lammers, S. M., Woods, R. J., Brotherson, S. E., Deal, J. E., & Platt, C. A. (2022). Explaining adherence to American academy of pediatrics screen time recommendations with caregiver awareness and parental motivation factors: mixed methods study. *JMIR Pediatrics and Parenting*, *5*(2), e29102.

Leung, C. Y., Hernandez, M. W., & Suskind, D. L. (2020). Enriching home language environment among families from low-SES backgrounds: A randomized controlled trial of a home visiting curriculum. *Early Childhood Research Quarterly*, *50*, 24-35.

Leung, C. Y., & Suskind, D. L. (2020). What parents know matters: Parental knowledge at birth predicts caregiving behaviors at 9 months. *The Journal of pediatrics*, *221*, 72-80.

Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge?. *Developmental psychology*, *46*(5), 1309.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Crown Currency.

List, J. A. (2024). "Optimally generate policy-based evidence before scaling," *Nature*, 626(7999), pages 491-499, February.

List, J.A. (2025). Experimental Economics: Theory and Practice. Chicago: University of Chicago Press.

List, J. A., Pernaudet, J., & Suskind, D. L. (2021). Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. *Nature communications*, *12*(1), 5765.

Luo, R., Song, L., Davis-Hilton, A., & Surrain, S. (2025). Parental beliefs and knowledge about dual language learning and education (KnDLLE) predict children's Spanish and English experiences at home. *Early Childhood Education Journal*, *53*(1), 287-298.

Lynch, J. (2017). Short-term professional learning that focuses on beliefs and practice: Improvements in preschool teachers' ongoing literacy practice. *Journal of early childhood teacher education*, *38*(2), 164-181.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research*, *35*(6), 382-386.

MacPhee, D. (1981). Knowledge of infant development inventory. *Unpublished manual and questionnaire. University of North Carolina, Chapel Hill*.

McCabe, A., Tamis-LeMonda, C. S., Bornstein, M. H., Brockmeyer Cates, C., Golinkoff, R., Wishard Guerra, A., ... & Song, L. (2013). Multilingual Children beyond Myths and toward Best Practices. Social Policy Report. Volume 27, Number 4. *Society for Research in Child Development*.

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early childhood research quarterly*, *28*(2), 314-324.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Pun, W. H., & Maczuga, S. (2019). Kindergarten children's executive functions predict their second-grade academic achievement and behavior. *Child development*, *90*(5), 1802-1816.

Mesman, J., Van Ijzendoorn, M. H., & Sagi-Schwartz, A. (2016). Cross-cultural patterns of attachment. *Handbook of attachment: Theory, research, and clinical applications*, *3*, 852-877.

Nocita, G., Perlman, M., McMullen, E., Falenchuk, O., Brunsek, A., Fletcher, B., ... & Shah, P. S. (2020). Early childhood specialization among ECEC educators and preschool children’s outcomes: A systematic review and meta-analysis. *Early Childhood Research Quarterly*, *53*, 185-207.

Noritz, G. H., Murphy, N. A., Neuromotor Screening Expert Panel, Murphy, N. A., Hagan Jr, J. F., Lipkin, P. H., ... & Yeargin-Allsopp, M. (2013). Motor delays: early identification and evaluation. *Pediatrics*, *131*(6), e2016-e2027.

Niklas, F., Cohrssen, C., & Tayler, C. (2016). Improving preschoolers’ numerical abilities by enhancing the home numeracy environment. *Early Education and Development*, *27*(3), 372-383.

O’Rear, C. D., Kirkland, P. K., & Purpura, D. J. (2024). The how many and give-N tasks: Conceptually distinct measures of the cardinality principle. *Early Childhood Research Quarterly*, *66*, 61-74.

Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*, *46*, 112-125.

Partika, A., Johnson, A. D., Phillips, D. A., Luk, G., Dericks, A., & Tulsa SEED Study Team. (2021). Dual language supports for dual language learners? Exploring preschool classroom instructional supports for DLLs' early learning outcomes. *Early Childhood Research Quarterly*, *56*, 124-138.

Planalp, E. M., O’Neill, M., & Braungart-Rieker, J. M. (2019). Parent mind-mindedness, sensitivity, and infant affect: Implications for attachment with mothers and fathers. *Infant Behavior and Development*, *57*, 101330.

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health*, *29*(5), 489-497.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, *30*(4), 459-467.

Purpura, D. J., Logan, J. A., Hassinger-Das, B., & Napoli, A. R. (2017). Why do early mathematics skills predict later reading? The role of mathematical language. *Developmental Psychology*, *53*(9), 1633-1642.

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review*, *45*(2), 250-267.

Rideout, V. (2021). The Common Sense Census: Media Use by Kids Age Zero to Eight in America, A Common Sense Media Research Study,[United States], 2013, 2017.

Rikhy, S., Tough, S., Trute, B., Benzies, K., Kehler, H., & Johnston, D. W. (2010). Gauging knowledge of developmental milestones among Albertan adults: a cross-sectional survey. *BMC public health*, *10*(1), 183.

Romeo, R. R., Leonard, J. A., Grotzinger, H. M., Robinson, S. T., Takada, M. E., Mackey, A. P., ... & Gabrieli, J. D. (2021). Neuroplasticity associated with changes in conversational turn-taking following a family-based intervention. *Developmental cognitive neuroscience*, *49*, 100967.

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological science*, *29*(5), 700-710.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development*, *83*(5), 1762-1774.

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist?. *Developmental disabilities research reviews*, *17*(2), 141-150.

Rowe, M. L., Denmark, N., Harden, B. J., & Stapleton, L. M. (2016). The role of parent education and parenting knowledge in children's language and literacy skills among White, Black, and Latino families. *Infant and Child Development*, *25*(2), 198-220.

Rowe, M. L., & Leech, K. A. (2019). A parent intervention with a growth mindset approach improves children's early gesture and vocabulary development. *Developmental science*, *22*(4), e12792.

Sawyer, B. E., Manz, P. H., & Martin, K. A. (2019). Supporting preschool dual language learners: Parents' and teachers' beliefs about language development and collaboration. In *Research in Young Children's Literacy and Language Development* (pp. 409-428). Routledge.

Scalise, N. R., Daubert, E. N., & Ramani, G. B. (2018). Narrowing the early mathematics gap: A play-based intervention to promote low-income preschoolers' number skills. *Journal of Numerical Cognition*, *3*(3), 559-581.

Schrank, F. A., & Wendling, B. J. (2018). The Woodcock--Johnson IV. *Contemporary intellectual assessment: Theories, tests, and issues*, *383*.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental science*, *11*(5), 655-661.

Sim, S. S., Berthelsen, D., Walker, S., Nicholson, J. M., & Fielding-Barnsley, R. (2014). A shared reading intervention with parents to enhance young children's early literacy skills. *Early child development and care*, *184*(11), 1531-1549.

Siu, A. L. (2015). Screening for speech and language delay and disorders in children aged 5 years or younger: US Preventive Services Task Force recommendation statement. *Pediatrics*, *136*(2), e474-e481.

Squires, J., Bricker, D., & Twombly, E. (2002). Ages & stages questionnaires: Social-emotional. *Baltimore: Brookes*.

Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development*, *47*, 82-95.

Supanitayanon, S., Trairatvorakul, P., & Chonchaiya, W. (2020). Screen media exposure in the first 2 years of life and preschool cognitive development: a longitudinal study. *Pediatric Research*, *88*(6), 894-902.

Suskind, D. (2025). Ai could rewire the next generation of children. https://thehill.com/opinion/technology/5373324-ai-threats-child-development/

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., ... & Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of child language*, *43*(2), 366-406.

Suskind, D. L., Leung, C. Y., Webber, R. J., Hundertmark, A. C., Leffel, K. R., Suskind, E., ... & Graf, E. (2018). Development of the survey of parent/provider expectations and knowledge (SPEAK). *First Language*, *38*(3), 312-331.

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness?. *Current Directions in Psychological Science*, *23*(2), 121-126.

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, *104*(46), 17954-17958.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Tian, J., Bennett-Pierre, G., Tavassolie, N., Newcombe, N. S., Weinraub, M., Hindman, A. H., ... & Gunderson, E. A. (2023). A growth mindset message leads parents to choose more challenging learning activities. *Journal of Intelligence*, *11*(10), 193.

Walsh, K., L'Estrange, L., Smith, R., Burr, T., & Williams, K. E. (2024). Translating neuroscience to early childhood education: A scoping review of neuroscience-based professional learning for early childhood educators. *Educational Research Review*, *45*, 100644.

Waters, E., Merrick, S., Treboux, D., Crowell, J., & Albersheim, L. (2000). Attachment security in infancy and early adulthood: A twenty-year longitudinal study. *Child development*, *71*(3), 684-689.

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.

Winter, L., Morawska, A., & Sanders, M. (2012). The Knowledge of Effective Parenting Scale (KEPS): A tool for public health approaches to universal parenting programs. *The journal of primary prevention*, *33*(2), 85-97.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive psychology*, *24*(2), 220-251.

Zajac, L., Raby, K. L., & Dozier, M. (2020). Sustained effects on attachment security in middle childhood: results from a randomized clinical trial of the Attachment and Biobehavioral Catch-up (ABC) intervention. *Journal of Child Psychology and Psychiatry*, *61*(4), 417-424.

Zambo, D. (2008). Childcare workers’ knowledge about the brain and developmentally appropriate practice. *Early Childhood Education Journal*, *35*(6), 571-577.

Zimmerman, F. J., & Christakis, D. A. (2005). Children's television viewing and cognitive outcomes: a longitudinal analysis of national data. *Archives of pediatrics & adolescent medicine*, *159*(7), 619-625.

Zinsser, K. M., Shewark, E. A., Denham, S. A., & Curby, T. W. (2014). A mixed-method examination of preschool teacher beliefs about social–emotional learning and relations to observed emotional support. *Infant and child development*, *23*(5), 471-493.

Zubler, J., & Whitaker, T. (2022). CDC's revised developmental milestone checklists. *American family physician*, *106*(4), 370-371.

Appendix

Demographics for Cognitive Interviewing Study

Age: 18-20 years: 10 (3.41%), 21-30 years: 107 (36.52%), 31-40 years: 68 (23.21%), 41-50 years: 43 (14.68%), 51-60 years: 24 (8.19%), over 60 years: 16 (5.46%), no response: 25 (8.53%)

Gender: Men: 115 (39.25%) Women: 167 (57.00%), Other: 7 (2.39%), no response: 4 (1.37%)

Primary Language: English: 227 (94.54%), Spanish: 2 (0.68%), Other: 14 (4.78%), no response: 0, (0.0%)

Participant education level: Some high school, no diploma: 4 (1.37%), GED (high school equivalency): 1 (0.34%), High school diploma: 12 (4.10%), Trade, technical, or vocational school: 3 (1.02%), Some college credit, but no degree: 35 (11.95%), Associates Degree: 8 (2.73%), Bachelor's Degree: 123 (41.98%), Master's Degree: 87 (29.69%), Doctorate or professional Degree: 19 (6.48%), no response: 1 (0.34%)

Participant education level (categories collapsed): Less than a Bachelor's degree: 63 (21.50%), Bachelor's degree or higher: 229 (78.16%), no response: 1 (0.34%)

Race: American Indian/Alaska Native: 4 (1.37%), Asian: 64 (21.84%), Black or African American: 66 (22.53%), Native Hawaiian/Other Pacific Islander: 1 (0.34%), White/European American: 134 (45.73%), Other: 17 (5.80%), No response: 2 (0.68%)

Ethnicity: Hispanic/Latino: 28 (9.56%), Not Hispanic/Latino: 253 (86.35%), No response: 7 (2.39%)

Income: Under $10,000: 14 (4.78%), $10,000-$20,000: 12 (4.10%), $20,000-$30,000: 16 (5.46%), $30,000-$40,000: 21 (7.17%), $40,000-$50,000: 18 (6.14%), $50,000-$75,000: 51

(17.41%), $75,000-$100,000: 48 (16.38%), $100,000-$200,000: 60 (20.48%), Over $200,000: 15 (5.12%), No response: 38 (12.97%)

Relationship Status: Divorced: 18 (6.14%), Living with partner: 41 (13.99%), Married: 95 (35.42%), Single: 119 (40.61%), Widowed: 5 (1.71%), Separated: 5 (1.71%), No response: 10 (3.41%)

Education/training in birth to 5: Training reported: 55 (18.77%), No training reported: 232 (79.18%), No response: 6 (1.05%)

Parent Status: Parents: 121 (41.30%), Non-Parents: 166 (56.66%), No response: 6 (2.05%)

Parent Gender (*% of all parents)*: Mothers: 65 (53.72%), Fathers: 55 (45.45%), No response (for gender): 1 (0.83%)

Parents' Number of Children *(% of all parents):* 1 child: 48 (39.67%), 2 children: 51 (42.15%), 3 children: 17 (14.05%), 4 or more children: 3 (2.48%), No response: 2 (1.65%)

Child Age (*% not reported because parents can fall into multiple categories*): Parents with a child 0-5 years: 58, Parents with a child 6-17 years: 45, Parents with a child 18 years or older: 40