Learning Variability Network Exchange (LEVANTE): A global framework for measuring children's learning variability through collaborative data sharing

Michael C. Frank (Stanford University), Heidi A. Baumgartner (Stanford University), Mika Braginsky (Stanford University), George Kachergis (Stanford University), Amy Lightbody (Stanford University), Robert Z. Sparks (Stanford University), Rebecca Zhu (Stanford University), Stephanie M. Carlson (University of Minnesota), Sandra Graham (University of California, Los Angeles), Sebastián J. Lipina (Universidad Nacional de las Artes, CEMEC-CONICET), Nora S. Newcombe (Temple University), Candice L. Odgers (University of California, Irvine), Robert C. Pianta (University of Virginia), Robert S. Siegler (Columbia University), Margaret Snowling (University of Oxford), Hirokazu Yoshikawa (NYU), Ana Cubillo* (Jacobs Foundation), Kenneth A. Dodge* (Duke University)

Author note: * indicates shared last authorship. This work was supported by a grant from the Jacobs Foundation. Please address correspondence to mcfrank@stanford.edu.

Abstract

Despite the ubiquity of variation in child development within individuals, across groups, and across tasks, timescales, and contexts, dominant methods in developmental science and education research still favor group averages, short snapshots of time, and single environments. The Learning Variability Network Exchange (LEVANTE) is a framework designed to enable coordinated data collection by research teams worldwide, with the goal of measuring variability in children's learning and development. The LEVANTE measure set aims to capture variability in learning outcomes (literacy and numeracy) as well as in core cognitive and social constructs. LEVANTE will yield a large, open access longitudinal dataset for long-term research use, both creating a multidisciplinary research network and facilitating the science of learning variability.

Introduction

This article introduces the Learning Variability Network Exchange (LEVANTE), a framework for developmental data collection and sharing with the aim of understanding the nature of developmental variability alongside its counterpart, developmental consistency. Our goal is to gather a large, rich, multi-context dataset with global scope that measures developmental change over time in children aged 2–12 years. Initiated by the Jacobs Foundation, LEVANTE is organized around a partnership between the Foundation, the data coordinating center (which develops the technical framework), an independent steering committee, and a network of global sites involved in data collection. Unlike classic cohort studies, LEVANTE provides a flexible framework in which funded sites as well as other researchers can measure variability in learning and development using consistent, state-of-the art tools for data collection and data management.

LEVANTE includes assessments of language and literacy; numeracy and math; social and emotional cognition; executive function; and spatial cognition and reasoning in children ages 2–12 years. These aspects of individual children will be examined in light of contextual constructs at the level of the self, the home and family, the school, and the community. The youngest children will be able to complete a subset of the direct assessments, allowing for continuity of measurement; the oldest will already be well along their academic trajectory, allowing measurement of learning variability. This developmental range also encompasses both the preschool period and the transition to formal education, allowing connections between constructs measured worldwide in early childhood and those studied in school-aged children. The overall goal of LEVANTE is to provide the research community with open tools and data so as to advance global research on variability and broaden participation in research. In the remainder of the introduction, we describe the theoretical rationale for LEVANTE and then discuss the idea of a federated cohort study in more detail.

Theoretical motivation

Variability – defined as the dispersion of measurements around some central tendency – is a common denominator in the study of learning and development. Individual children's abilities vary with age, but also vary across different tasks and even within the same task at multiple timescales (Kulke et al. 2018; Nussenbaum et al., 2022; te Brinke et al., 2023). Within classrooms, children of similar biological age show varied skills (Carstensen et al., 2019; Hughes & Devine, 2015; Mabbott & Bisanz, 2003; Nelson, 1981; Smiley & Dweck, 1994). Yet currently dominant methods tend to favor group averages, short snapshots of time, and single environments. When variability is acknowledged, researchers disagree about its origins, how to address it, and when to embrace vs. when to reduce it (Ellis et al., 2022; Frankenhuis, Young, & Ellis, 2020; Frankenhuis & Nettle, 2019; Kievit et al., 2013). Similarly, there is a lack of data and scientific evidence on the interplay among different dimensions of variability – for example, about the relation between day-to-day variation in cognition and developmental change over the course of several years.

Studying variability could facilitate our understanding of why variation might be especially prevalent in specific groups or contextual circumstances (Jacobs Foundation, 2023). Identifying the sources of variability is essential not only for mapping the limits of empirical findings but also to generate a comprehensive theory of children's development. That is, if the relation between an environmental experience and a child's learning is reliably modified by environmental context, our theories of how experience shapes development will be enriched. For example, Lansford and colleagues (2005) found that parental harsh discipline of a child predicted growth in that child's externalizing behavior problems, but the strength of the relation was diminished under cultural circumstances in which harsh discipline was more normative, suggesting that the culturally-informed meaning of the parental behavior influences the effects of the behavior. In sum, variability – and its opposite, consistency – can provide important clues to mechanisms of learning and development and how they operate across contexts (Frank et al., 2021).

Additionally, researchers might consider investigating individual variability to promote more effective customized learning interventions. For example, if students with reading difficulties show different latent profiles of language and cognitive skills (Kulesz et al. 2024), researchers and educators could potentially improve reading outcomes more effectively by providing customized interventions based on these profiles (Connor et al., 2007). If research fully embraced learning variability, the resulting knowledge could enable policy makers, designers, and educators to make decisions that more effectively serve a greater variety of children.

Finally, there is a significant gap in developmental psychology research involving children outside of Western convenience sample contexts (Henrich et al., 2010; Kidd & Garcia, 2022; Nielsen et al., 2017; Singh et al., 2023; Nag et al., 2024). If global samples are not included in research, it is impossible to quantify variability across cultures and contexts. This gap has important consequences both practically and theoretically. Practically, targeting interventions to new populations is risky without understanding how contexts vary (Bryan et al., 2021). And theoretically, claims about the universal ingredients of development are impossible without measurements of variation across the global population (Frank et al., 2021). Thus, diversifying developmental psychology to better understand individual and contextual variation is thus both a major challenge and a significant need.

A federated cohort study

Studying variability and diversity requires large datasets. Post-hoc data harmonization (e.g., gathering and coordinating data from multiple groups post-data collection) can yield valuable resources, but this process is labor-intensive and relies on the existence of pre-existing datasets with the desired characteristics (Frank et al., 2017; Zettersten et al., 2023; Gilmore et al., 2016). For research questions that require multiple measures, large-scale cohort studies are often the only option; yet new innovations in measurement can lead to the need for new datasets to be created.

Cohort studies have been a critical part of the history of developmental research (e.g., Roche, 1992). They offer special insight into variability by providing measures of change over time. But such studies are immense undertakings, spanning over many years and requiring extensive investments of time and money. For example, the Adolescent Brain and Cognitive Development study (ABCD) is projected to run for more than 15 years and to cost many hundreds of millions of dollars. In the ABCD protocol, a group of US sites pursues coordinated data collection following a cohort of adolescents longitudinally (Volkow et al., 2018). While ABCD is among the most comprehensive, many other such cohort studies both within and outside of the US – such as the Environmental Risk Longitudinal Twin Study (Fisher et al., 2015), Fragile Families and Child Wellbeing Study (Waldfogel et al., 2010), Texas Twin Project (Harden et al., 2013), and Twins' Early Development Study (Oliver & Plomin, 2007) – provide important resources for developmental scientists. However, datasets that provide deep characterization of children's minds, brains, and genes are typically limited to one geographic area or country (e.g., Golding et al., 2001).

Coordinated data collection of the type exemplified by cohort studies provides important scientific opportunities, but also carries a number of costs. First, each site participating in a cohort study must recruit and collect data according to the study protocol, with no opportunity to harmonize the data collection with ongoing efforts (e.g., existing local cohorts). Second, there is typically at best limited opportunity for sites to add measures in order to address site-specific questions. Third, this substantial resource investment typically requires that each site have specific capacities and characteristics, often limiting opportunities for global participation (cf. Lansford & Bornstein, 2011). Here, in contrast, we explore a more flexible, federated approach to the longitudinal cohort study, in which a harmonized dataset emerges from coordinated, distributed data collection across a set of sites pursuing their own research goals using a common platform.

While traditional cohort studies either focus on a single site or else mandate identical, coordinated data collection across sites, LEVANTE data collection will be distributed across sites via calls for proposals that allow participating researchers to integrate core measures into new or existing study designs. These core measures are designed from the ground up to provide an openly accessible, internationalized set of tasks and surveys created with psychometric best practices in mind. LEVANTE will result in an interrelated set of longitudinal studies that use the same core measure set and similar sampling plans to explore developmental variability within and across diverse populations. The product will be a high-value dataset that enhances our understanding of variability in human development across diverse contexts.

LEVANTE embraces the principles of open science as a fundamental design feature (Klein et al., 2018; Frank et al., 2024). All measures designed for LEVANTE will be open and permissively licensed, allowing their adoption and reuse throughout the developmental research community at no cost to users. To accelerate collaborative investigation of the main dataset, we aim to release data on the core study measures soon after they are collected, rather than waiting for individual sites to terminate data collection. The dataset will be designed to preserve privacy through extensive de-identification, allowing free and permissive sharing of the resulting data. The overall goal is to create a set of research products – data, code, tasks, and materials – that together accelerate progress in the developmental and learning sciences.

The current manuscript

The aim of the current manuscript is to describe the design of the LEVANTE framework. We begin by presenting our general plans for data collection and our measures. We then describe the scientific aims of the project and broader considerations around ethics, privacy, and data use. We end by considering LEVANTE in the context of our pilot data collection in three geographically, culturally, and linguistically diverse sites – Ontario, Canada; Bogota, Colombia; and Leipzig, Germany – and by outlining potential benefits of the LEVANTE approach for researchers, teams, and the field as a whole.

The LEVANTE Framework

The key feature of LEVANTE is the federated data collection model, in which participating sites collect data from their local population using a shared set of measures. LEVANTE projects are expected to be diverse, involving the collection of LEVANTE measures in new samples, the addition of LEVANTE measures to ongoing or planned studies, or the supplementation of LEVANTE data collection with other measure types.

We anticipate enrolling a series of cohorts, with start dates for data collection ranging from 2025 to 2030. A major goal of LEVANTE site selection is to recruit sites that sample from communities that are traditionally under-represented in research. To facilitate longitudinal data analysis and analysis of within-person variability, sites will be required to collect data at a minimum of three timepoints per child.

The first research wave of the program will focus on enrolling children ages 5 - 10 in a small set of target languages including English, Spanish, and German as well as others as determined based on the availability of measures and the key learning questions outlined in the Jacobs Foundation Research Agenda (Jacobs Foundation, 2023). This decision reflects the relative maturity of measures designed for older children (see below). The second research wave will focus on extension of the LEVANTE framework to early childhood (enrollments at 2 - 4 years). Subsequent research waves will focus on extension of the LEVANTE framework to broader geographic, cultural, and linguistic contexts. Additional focus areas for future waves may include measurement of educational context; inclusion of neuroscience measures; or dense, within-person measurement for modeling and predictive validity.

Each LEVANTE site will recruit participants via one of two standard paths: school-based recruitment – where families in the same school are recruited – or family-based recruitment – where families are recruited individually (e.g., via online advertising, a community-based setting, or other recruitment method). Younger children will require adults (parents, teachers, or research staff) to oversee testing on a one-to-one basis. In contrast, older children may be able to navigate administration of the measures on a computer or tablet platform in a group setting (for example in a classroom). In all cases, children and caregivers will contribute data; in cases of school-based recruitment, we will additionally ask sites to administer questionnaires to teachers to better characterize children's classroom learning environments.

For each of three annual assessments, we anticipate a total assessment burden for the core measures of approximately one hour for children aged 5 - 12 years (sometimes broken into multiple sessions). Caregiver questionnaires will take no more than one hour, and teacher questionnaires will take no more than 30 minutes. This relatively low assessment burden will allow the integration of LEVANTE assessments with other measures and with ongoing or planned research.

Sites will make use of a shared technical infrastructure (Figure 1). Each site will access a dashboard allowing them to administer questionnaires and direct assessments in a web browser or on a tablet. Data will be transferred from this administration platform to the LEVANTE data repository via a data validation interface and disseminated in combination with de-identified demographic data. Sites will have immediate access to the data they collect, but all data collected on the LEVANTE core measures will be released publicly on a twice-yearly schedule following a six month embargo period, allowing quick access to data.

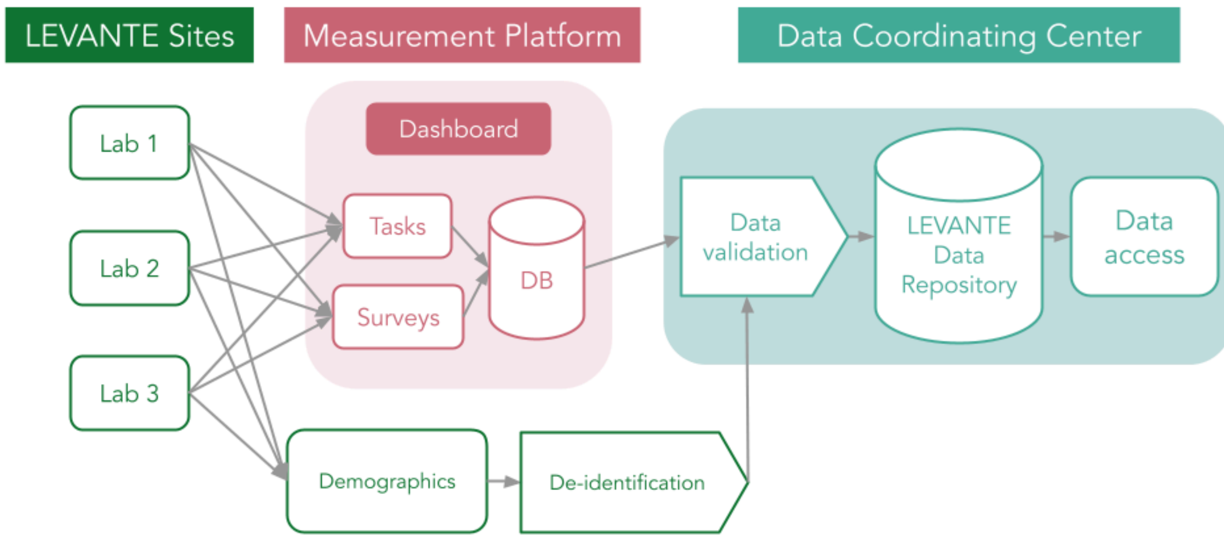

Figure 1. Data flows from partner sites to the data coordinating center and LEVANTE data repository.

Measurement and Sampling

Measure selection

The LEVANTE core assessment battery is designed to provide holistic measurement of learning and development for children ages 2 – 12 years. Figure 2 shows a map of the broad construct areas that we aim to assess. However, there are fewer well-validated measures for the youngest children; thus LEVANTE will begin with children ages 5 –12 and engage in iterative piloting of measures for children ages 2 – 4. The current manuscript reports only on measures for older children. We used the following set of principles to guide measure selection for

LEVANTE. While few measures fulfill all criteria, the balance of these factors was imperative to the decision process. We selected measures based on the following desiderata:

- **Short.** Given the pressure to assess multiple constructs in a short timeframe, measures needed to be brief.
- **Reliable.** Since measurement of variability is the key goal, we prioritized measures with demonstrated high test-retest reliability and/or internal consistency as demonstrated in the literature.
- **Valid.** Given that we aim to generate evidence that could have translational impact on different education sectors, we selected measures with evidence of strong construct, predictive, and external validity.
- **Cross-culturally appropriate.** Since measuring children's variability across contexts is a key goal, we sought measures with demonstrated use in a wide range of cultures and contexts, including low- and middle-income countries (LMIC).
- **Broad age range.** Since measuring within-children's variability and developmental growth across contexts are goals for LEVANTE, we looked for measures that could be used across our full age range.
- **Normed.** Since measuring children's variability across ages, groups, and contexts is a key goal for LEVANTE, we prioritized pre-existing measures with norms across our age range of interest.
- **Accessible and non-commercial.** Because we intend the LEVANTE measures to be freely available, we only considered including measures that could be used without incurring a licensing cost.

Our overriding impulse in measure selection for LEVANTE was pragmatism: with a short timeline and a very limited budget of time for each wave of assessment, we needed to be aggressive in selecting a relatively small number of constructs, a small number of measures associated with each construct, and a limited number of items for each construct. The measure selection process entailed an extensive examination of the literature, conversations with experts in the field (i.e., individual faculty members as well as a faculty advisory group with expertise in diverse areas of child development and education), and a cost-benefit analysis considering factors that included length of assessment, reliability and validity across contexts, and developmental applicability.

Child assessments

Our core constructs and subconstructs for direct assessment (Table 1) were selected based on an interest in examining learning outcomes and their precursors in early childhood, combined with the goal of creating a holistic assessment of individual children. The direct assessments we selected for LEVANTE have been instantiated with many different sets of parameters without consensus as to a single standard implementation. Even in the case of well-known measures, length constraints made using some tasks infeasible without modification. Thus, for many of these, we created what is essentially a new instrument, albeit one derived from the previous literature and in some cases making use of previously published stimuli.

From the perspective of learning outcomes, we identified language/literacy and numeracy/mathematics as key outcome domains in which we could track continuity between early childhood and later educational outcomes (Marchman & Fernald, 2008; Schneider et al., 2018). To these constructs, we added reasoning, executive function, spatial cognition, and social cognition as key domains of interest for both cognitive development and education. Measures of reasoning – for example, matrix reasoning tasks – are highly correlated with educational outcomes (Downey et al., 2014; Green et al., 2017; Pind et al., 2003) and provide a widely used tool for the reliable characterization of individual variation in cognition. Executive function and self regulation have similarly been linked to educational and life success (e.g., Ahmed et al., 2019; Moffitt et al., 2011). Spatial reasoning is linked to STEM outcomes (Atit et al., 2022; Tian et al., 2023). Finally, social cognition is an important area of interest in early childhood, especially with respect to cross-cultural (e.g., Callaghan et al., 2011) and socioeconomic variation (Fendinger et al., 2023). To assess key attitudes, we also ask children a small number of questions about their well-being, peer relationships, and their feelings about schooling. In sum, our direct assessment constructs include: language & literacy, numeracy & mathematics, reasoning, executive function, social cognition, and spatial cognition.

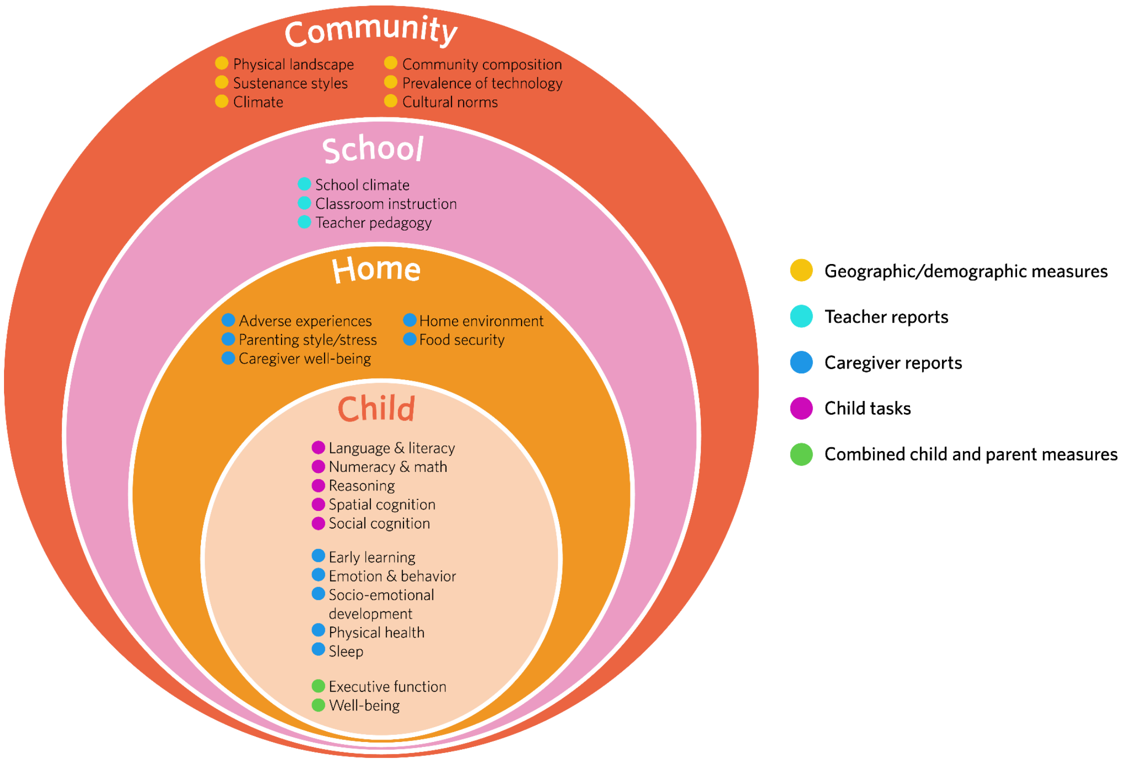

Figure 2. Visualization of child, home, school, and community constructs to be measured by LEVANTE data collection efforts.

Construct	Subconstruct	Task Name	Task Description	Citation
Language & Literacy	Vocabulary	ROAR-Vocab	Match a word to a picture	Yeatman et al., 2021
	Phonological Awareness	ROAR-PA	Select the word that starts with the same letter as the target word	Yeatman et al., 2021
	Word Recognition	ROAR-WR	Identify words and non-words (i.e., lexical decision task)	Yeatman et al., 2021
	Sentence Reading	ROAR-SRE	Identify whether sentences are true or false	Yeatman et al., 2021
	Grammar	TROG	Match a phrase to a picture	Bishop (1983)
Number & Math	Formal Math	Early Grade Mathematics Assessment (EGMA)	Answer addition and subtraction questions	Platas et al., 2016
	Early Numeracy	Number Line Estimation	Place a given number on a line from 0-10 or 0-100	Schneider et al., 2018

Reasoning	Pattern Recognition	Matrix Reasoning	Select pattern that completes 3x3 matrix	Raven, 2000
Executive Function	Working Memory	Dot Matrix	Tap a sequence of dots in a prescribed order	Arce & McMullen, 2021
	Response Inhibition	Hearts and Flowers	Select right or left key based on given rule	Camerota et al., 2020
	Set Shifting	Same Different Selection	Match picture to another picture	Obradović et al., 2025
Social Cognition	Social Cognition	Theory of Mind Battery	Listen to a short vignette and answer questions about the vignette	Sotomayor-Enriquez et al., 2024
	Hostile Attribution Bias	Hostile Attribution Bias Subscale of the Social Information Processing–Attribution Bias Questionnaire	Answer questions about ambiguous social situations	Dodge et al., 2015
Spatial Cognition	Mental Rotation	Mental Rotation Task	Match silhouette to rotated picture	Shepard and Metzler, 1971

Table 1. List of child direct assessment constructs and measures.

Caregiver reports

Primary caregivers can provide a wealth of information about their children and the broader contexts of influence on their children's development (Table 2). All caregivers will provide a set of basic demographic variables including indicators of community membership and socioeconomic status (Singh et al., 2024). In addition to demographics, we survey constructs relevant to primary caregivers' involvement and the child's environment, specifically from the primary caregiver's view of the child (Bourdon et al., 2005; Goodman, 1997; 2001; Janitza et al., 2020; Murphey, 1992) and the primary caregiver's view of the child's home and school environments (Davis-Kean, 2005; Lansford et al., 2023; Matheny et al., 1995; Okagaki et al., 1998; Seefeldt et al., 1998). Questions about the home environment will similarly probe variables relating to educational achievement, including both aspects of the caregivers' parenting style (Dornbusch et al., 1987; Essau et al., 2006; Frick et al., 1999; Shelton et al., 1996; Zimet et al., 1990) as well as broader features of the home environment including food security (Call et al., 2024; Heflin et al., 2022), structure (Garrett-Peters et al., 2016; Matheny et al., 1995; Micalizzi et al., 2019), and literacy/numeracy practices (Lansford et al., 2023; Manolitsis et al., 2013; Napoli & Purpura, 2018).

Questions about the child focus on key variables relating to educational attainment, such as their health and well-being (Cardenas et al, 2022; Essex et al., 2002; Koita et al., 2018; Ye et al., 2023), their socioemotional development (Hammer et al., 2018; Pettit et al., 1991) and their executive function and self-regulation (Ahmed et al., 2019; Bourdon et al., 2005; Moffitt et al., 2011). These reports from caregivers will add new information about the child in addition to providing a complementary picture to overlapping direct child assessment tasks such as social cognition and executive function.

Target	Construct	Subconstruct	Source	Citation
Caregiver	Child Health/ Well-Being	General Demographics	ManyBabies Demographics	Singh et al., 2024
		General Health	National Survey of Children's Health	NSCH, 2022
		Cognition/Learning	National Survey of Children's Health	NSCH, 2022
		Sleep Problems/Health	Patient Reported Outcomes Measurement Information System (PROMIS)	Forrest et al., 2018
		Sex/Gender	Adapted from "Measuring Sex, Gender and Sexual Orientation" Report	National Academies of Sciences, Engineering, and Medicine, 2022
		Puberty	Youth Pubertal Development Scale	Peterson et al., 1988
		Adverse Events	Pediatric ACEs and Related Life Events Screener (PEARLS)	Aces Aware, 2024
		Technology Use	Developed by LEVANTE researchers	

		Executive Function	Developed by the Global Executive Function Initiative	
	Social Emotional	Emotion Problems/Regulation	Social Competence Scale	CPPRG, 1995
		Peer Relationships	National Survey of Children's Health	NSCH, 2022
		Conduct/Behavior	Child Behavior Questionnaire, Social Competence Scale, Jukes Social Emotional Competencies	CPPRG, 1995; Jukes et al., 2021
		Prosocial Behavior/Communication	Child Behavior Questionnaire, Social Competence Scale, Jukes Social Emotional Competencies	CPPRG, 1995; Jukes et al., 2021
		Curiosity	Jukes Social Emotional Competencies	Jukes et al., 2021
		Academic Interest	National Survey of Children's Health	NSCH, 2022
	Home Environment	Early Learning	Trends in International Mathematics and Science Study Early Learning Survey	TIMSS, 2018
		Community Environment	National Survey of Children's Health	NSCH, 2022
		Community Safety	National Survey of Children's Health	NSCH, 2022
		Food Security	USDA Food Security Module Short Form	Bickel et al., 2000
		Learning Materials/ Opportunities	Home Observation for Measurement of the Environment (HOME-21)	Lansford et al., 2023
		Family Companionship	Home Observation for Measurement of the Environment (HOME-21)	Lansford et al., 2023
		Encouragement of Maturity	Home Observation for Measurement of the Environment (HOME-21)	Lansford et al., 2023
		Structure	Chaos, Order, and Hubbub Scale (CHAOS)	Matheny et al., 1995
	Parenting	Discipline	Multiple Indicator Cluster Survey	UNICEF, 2024
		Stress	National Survey of Children's Health	NSCH, 2022
		Warmth	Rohner Parental Acceptance-Rejection/Control Questionnaire	Rohner, 2005
		Control	Rohner Parental Acceptance-Rejection/Control Questionnaire	Rohner, 2005
		Social Support	Multidimensional Scale of Perceived Social Support (MSPSS)	Zimet et al., 1988
	Caregiver Well-Being	Anxiety	Patient Health Questionnaire-4	Kroenke et al., 2009
		Depression	Patient Health Questionnaire-4	Kroenke et al., 2009
		Discrimination	Everyday Discrimination Scale	Williams et al., 1997
		Life Changes	Recent Life Changes Questionnaire	Holmes & Rahe, 1967

		Social Status	MacArthur Scale of Subjective Social Status - Adult Version	Adler et al., 2000
Teacher	Teacher Demographics	Background	Comprehensive Teacher Survey	ECLS-K, 2000
	Classroom	Characteristics	Comprehensive Teacher Survey, ECLS Teacher Questionnaire	ECLS-K, 2000
		Instruction	ECLS Teacher Questionnaire	ECLS-K, 2000
	School Climate	Teacher Belonging	ECLS Teacher Questionnaire	ECLS-K, 2000
		Teacher and Student Safety	ECLS Teacher Questionnaire	ECLS-K, 2000
	Pedagogy	Feelings about Job	Teacher Beliefs and Experiences Scale	Colaner, 2016
		Beliefs about Teaching	Teacher Beliefs and Experiences Scale	Maslach & Jackson, 1981
		Ideas about Children	Teacher Beliefs and Experiences Scale	Tschannen-Moran & Woolfolk Hoy, 2001
	Family/Students	Connection and Communication	ECLS Teacher Questionnaire	ECLS-K, 2000
Child	Teacher	Student-Teacher Relationship	Child Interview 5-6	Ruzek et al., 2020
	Peer	Peer Relationships	Early Childhood Longitudinal Study	ECLS-K, 2004
	School Climate	Belonging	Child Interview 5-6	Ruzek et al., 2020
		Safety	LEVANTE	LEVANTE, 2024
	Academics	Academic Perception	Early Childhood Longitudinal Study, Child Interview 5-6	ECLS-K, 2004
		Growth Mindset	Child Interview 5-6	Ruzek et al., 2020

Table 2. List of parent report constructs and measures.

Other measures

LEVANTE aims to characterize the home environment of individual children as well as other meaningful contexts, including their school and geographic (neighborhood) environments. Each participating LEVANTE data collection site will provide meta-data about the specifics of their data collection site. These meta-data will be linked to collected data about the specifics of their data collection site. These details will include informing how sampling relates to the broader population of the site location in terms of sociodemographic characteristics. Sites will also be asked to provide other meta-data for the population from which they are recruiting, including physical landscape, climate, community composition, salient cultural norms around parenting and childhood, typical structure of schooling, and prevalence of technology.

Both school and neighborhood environments account for some of the variation in early childhood development (Minh et al., 2017). To measure environmental variables, individual contributing partners will use identifiable location data to derive a set of de-identified geographic features for each child's primary household location (and in cases of school-based administration, school location). These will include two sub-constructs: neighborhood built environment measures and general environmental measures. For built environment measures, availability of variables may change by site location, but we anticipate access to population density (rural/urban classification), greenspace, poverty, local area inequality (GINI coefficient), and walkability score. For general environmental measures, we anticipate accessing average temperature, temperature at date of administration, heat index, daylight, night-time light pollution, noise pollution, and air pollution metrics (see e.g., Kühn & Gallinat, 2024).

For sites that recruit families through schools, we will provide a short questionnaire for teachers that captures both classroom and broader school context, including constructs such as school/classroom climate, classroom composition, pedagogical attitudes, teacher satisfaction, administrative support, and school/classroom resources (Table 2).

Measure adaptation

We created the initial versions of the LEVANTE measures with substantial input from local researchers at each of the three pilot sites (see below). At time of writing we are engaged in iterative piloting and development of these tasks, in light of psychometric analyses of pilot data.

New sites will join the framework and bring with them new contexts and new languages. We do not expect that the LEVANTE measures will be appropriate for every context; outside of contexts in which formal schooling is common, it will likely be inappropriate to apply measures of the type we have created (Greenfield, 1997). Even in contexts where the LEVANTE measures are appropriate, some adaptation will be necessary. To take an obvious example, the grammar measure will need to be adjusted for each new language; other measures will also almost certainly need customization as well. To create customized, contextually-sensitive measures, new LEVANTE sites will be able to make modifications of the current measures with support from the Data Coordinating Center. They can then choose to either validate these new measure variants by collecting a separate validation sample or to use their new, unvalidated measure variants for the first wave of their main study and then adjust these measures for subsequent waves based on psychometric analyses of first wave data.

Scientific Aims

LEVANTE is designed to help researchers, educators, and policy-makers understand variability in child development and learning across individuals, groups, and contexts, thus improving future outcomes for diverse groups of children worldwide. A set of key scientific aims follow from this purpose. These initial aims will be pursued through analytic collaborations among the principals of LEVANTE (e.g., the data coordinating center, the steering committee, and the network of participating sites), but the set of opportunities afforded by the LEVANTE dataset is vast and the intention is that it will be reused by many groups to answer questions related to these general aims.

Exploiting the longitudinal structure of the dataset, LEVANTE's primary analytic approach is to measure growth over time within individuals and within sites, and to consider how variation in growth is influenced by contextual factors. Based on the observational nature of the dataset, these initial aims are descriptive, rather than causal, but description of the structure of variation in learning and development itself has substantial value, especially when pursued on a global scale. Each of these analyses will be developed during the initial stages of the project. Because of the complexity of the modeling efforts, we assume that preregistration of analysis plans completely *a priori* will not be possible. Instead, we intend to develop analytic models using subsets of the data (including pilot data and initial longitudinal measures) and then preregister these models for application to the growing dataset in collaboration with interested researchers.

Our first scientific goal is to estimate how children's growth trajectories are influenced by variation in the overlapping contexts of their development (including their home, school, and neighborhood). For example, considering literacy, we can estimate the average trajectory for reading outcomes as well the influence of the home literacy environment on these outcomes (Rodriguez & Tamis-LeMonda, 2011; Schmitt et al., 2011). The size of this influence can in turn be compared across contexts. For example, we might find that the influence of the home environment is relatively larger in some contexts than others; or perhaps this relationship is consistent across contexts. Similar analyses can be conducted taking into account important sociodemographic moderators as well, for example indicators of socioeconomic status.

Our second scientific goal is to explore the dimensional structure of development. We can examine the dimensionality of individual constructs, such as reading (Lervåg et al., 2009; Tomblin & Zhang, 2006) or math (Milburn et al., 2019). Moreover, since LEVANTE initiatives involve a large number of constructs, sequenced over time, on wide-ranging samples, the dataset will yield in-depth knowledge regarding how different constructs might develop and interact in early childhood. For example, LEVANTE may help researchers investigate the relation between early emerging capacities (e.g., spatial cognition) and later emerging capacities (e.g., mathematical knowledge) in greater detail, across a wide range of contexts.

The third scientific goal is to go beyond characterization of average development to describe the structure of developmental variation. For example, if we are modeling the relation between two traditional constructs, (e.g., whether different measures of language and reading relate to one another), we can also examine variation in this construct structure across sites, as well as how variation in the construct structure is related to particular site-level moderators. The dimensional structure of development has been a persistent theoretical question in cognitive development (Breit et al., 2020; Hartung et al., 2018; Juan-Espinosa et al., 2006): how does the space of variation change across the lifespan? We will examine the factor structure explaining individual variation in the LEVANTE data (e.g., how many factors are required to explain variation across tasks) and test whether these factors are consistent across contexts and age.

Our fourth and final goal is measuring the impact of within-individual variability on learning outcomes. Recent work suggests an important role for within-individual variability in predicting individuals' preparedness to learn (Schmiedek et al., 2020). Toward this goal, we aim to estimate within-individual variability both within and across tasks, which requires longitudinal data with multiple measurements from a single child. In addition, we will use longitudinal models to estimate adherence to or deviation from predicted developmental trajectories, and investigate the relation between this variability and various learning outcomes. The availability of longitudinal data also provides the opportunity to investigate temporal relationships (e.g., to better understand the nature of the relationship between change in children's formal math and socio-emotional difficulties; Dobbs et al., 2006).

Given the design of LEVANTE, direct comparisons of outcomes between sites are not appropriate. Myriad differences in participant sampling and recruitment as well as in the particular circumstances in which measures are administered confound any inference about differences between sites. For example, while participants in one pilot site were recruited through their schools and tested on tablets in their classroom, participants in another site volunteered to participate through an online database and were tested on a heterogeneous mix of computers and tablets in their homes. Further, the research literature generally does not find evidence for metric or scalar invariance for child learning measures across countries and cultures (Asil & Brown, 2015). However, configural invariance is found much more commonly, suggesting that comparison of associations across countries and sites will be possible. Thus, LEVANTE is not designed to allow statements of the type "[site/country/culture] X scored higher than Y on measure M" and such statements should not be made on the basis of LEVANTE data. Instead, the utility of the data will be in performing parallel analyses across sites and comparing derived parameters from these sites, allowing for statements of the type "[sites/countries/cultures] X and Y both showed a positive relation between variables A and B although the strength of the relationship varied as a function of environmental context Z."

Ethics, Privacy, and Data Use

Rapid sharing of longitudinal data from children poses a number of potential risks. How should these risks be managed? The range of data use policies for "open" repositories in the developmental field is quite large, ranging from almost complete openness to relatively restrictive data use agreements. For example, CHILDES, the leading repository in the field of child language, makes data openly available for download with only minimal restrictions (a creative commons CC-BY-NC-SA license – which allows attributed, non-commercial redistribution of the dataset; MacWhinney, 2000); in contrast, the Adolescent Brain and Cognitive Development (ABCD) study requires an institutional agreement for data access, limiting reuse to academic investigators who can go through the – sometimes lengthy – process of finding an institutional signer who is willing to approve an agreement (Volkow et al., 2018).

More restrictive data sharing policies can mitigate privacy and legal risks effectively, but they come at a significant cost. The requirement of a signed institutional agreement is onerous even within the US academic framework, where it can take weeks or months to identify a university official who has the authority and time to sign a data use certification, and even more difficult in the international context. Within this landscape, the goal of the LEVANTE policies is to establish a coherent yet efficient framework for dataset creation that mitigates these risks. Here we enumerate a set of principles that guide our decision-making.

Efficiency of data access broadens global access and increases reproducibility. A primary guiding principle for LEVANTE is that the more efficiently we make data available, the more we will accelerate discovery with respect to our scientific aims. Unless they mitigate specific known risks, barriers to data access undermine the project's goal to provide measures and data that deepen our understanding of learning variability. Further, the more efficiently accessible LEVANTE data are, the more we mitigate scientific risks around research reproducibility. For example, if a particular LEVANTE data release is cited in a paper making a controversial claim, then the accessibility of this release to independent analysts will play a key part in allowing verification and investigation of that claim. Thus, openness can mitigate risks as well as creating them.

Independence of individual sites increases compliance and flexibility for global partners. In many consortium studies, all partners are required to follow a single ethics review process. In the global context, however, this kind of uniformity would be impossible for some research teams and would create substantial barriers for others. Recognizing these obstacles, global consortia such as *ManyBabies* have instead used a federated system in which individual contributors must document their own ethics review process and ensure that their work is compliant with local regulatory frameworks (Frank et al., 2017). We adopt this federated framework in LEVANTE so as to allow for ethical data collection and sharing while minimizing the cost of centralization.

De-identification of centralized data storage mitigates legal/regulatory risks. Sharing of identifiable participant data is complex from a regulatory perspective and requires substantial oversight. In contrast, sharing of fully de-identified data poses far fewer obstacles for contributors. Thus, the core LEVANTE dataset (composed of data collected on the LEVANTE measures) will be completely de-identified during all parts of the data collection and data sharing process, including an analysis of statistical reidentification risk and data blurring as necessary to mitigate this risk. This full de-identification policy mitigates privacy risk because only local investigators will ever have access to key identifying information (e.g., participant names, birthdates, etc.). It also ensures that data transfer is not governed under the prevailing legal frameworks (e.g., HIPAA and GDPR), meaning that research partners should be able to share data without creating specific research reliance agreements, which would create substantial legal/bureaucratic overhead.

Transparency mitigates scientific and reputational risks. For any scientific effort to share data, there is always the risk that claims made from that dataset will contradict the beliefs of the data creators. Cross-national datasets present specific risks in this regard in that they afford analysis of differences between nations and between demographic groups. Some of these analyses may be valid, whereas others may be undermined by differences in sampling, recruitment, or administration across groups. Both our scientific publications and the data repository itself will prominently describe our position on the (in-)validity of group level analyses.

Preregistration mitigates professional risks from data sharing. Since participating sites will be sharing data from core measures on an accelerated scale, the risk exists that analyses of some or all of a particular partner's dataset will be conducted by independent analysts (or other LEVANTE team members) prior to that partner's own analysis (i.e., the partner will be "scooped"). (Note that if site partners collected additional data beyond the core measures, they will not be required to deposit additional measure data until one year after the end of their LEVANTE grant, providing additional publication opportunities for data they collect beyond the core measure set.) For sites to protect their publication opportunities and to preserve the integrity of scientific data analyses, we encourage them to adopt the following practices. First, for key research questions, partners should consider submitting registered reports on their key questions prior to data collection if possible, or at least prior to data analysis. Registered reports are now accepted by many leading developmental psychology journals (e.g., Roisman et al., 2023). Second, we suggest that partners preregister their key hypotheses. Finally, we suggest that partners who are concerned about precedence risks integrate their hypotheses around key measures with analyses of other variables not included in the core LEVANTE measure set. Not only will this group of practices mitigate "scooping" risks, they will also decrease the chance of inflated effect estimates due to post-hoc data exploration and selective publication.

LEVANTE will encourage appropriate citation practices. We will encourage LEVANTE sites to write a short paper describing their data collection and initial results in order to provide a target for citations of uses of their dataset. Furthermore, sites will provide a list of ongoing projects and analyses using their data, with links to registered reports or other preregistrations if available, and a contact address. Dataset users will be encouraged to review this list and reach out prior to working on pre-planned analyses. These steps will also decrease scooping risks for external data users, who may not know if a site partner is already in process of publishing an analysis that the external data user hopes to complete.

Governance

LEVANTE has a governance structure that encompasses the Jacobs Foundation and its leadership, an external steering committee, and a Data Coordinating Center. The Jacobs Foundation is committed to supporting research on learning variability in children between the ages of 2 and 12 in high-, middle- and low-income countries, through grantmaking and other research programming. With advisory input from the steering committee and external reviewers, the Jacobs Foundation selects and funds the research sites involved in LEVANTE.

The LEVANTE Steering Committee is a group of developmental scientists who are leaders in their respective fields, including individuals from a range of disciplines, regions/cultural backgrounds, and genders. Their main responsibility is to provide oversight over the overall LEVANTE framework, and to monitor its progress. Furthermore, their duties include strategizing, designing, and implementing the program, supervising measure selection and the sampling plan, specifying scientific criteria to assess applications, as well as approving proposed priorities for research partner proposals. The Steering Committee is appointed by the Jacobs Foundation and members serve three year terms; members are excluded from consideration of grant awards.

The Data Coordinating Center (DCC) at Stanford University is charged with the creation, implementation, and maintenance of the measurement and data storage frameworks developed for LEVANTE. The DCC develops the measures included in LEVANTE. It further manages the design, building, and execution of the infrastructure for LEVANTE data collection, storage, and access.

Pilot Studies and Future Samples

Developing a set of reliable, valid, and efficient measures requires extensive piloting. We identified pilot sites in Ontario, Canada; Leipzig, Germany; and Bogota, Colombia to collect cross-sectional data on the full set of LEVANTE measures. These sites were selected through a combination of the Jacobs Foundation's programmatic objectives, the desire to pilot remote, lab-based, and school based administration, and pre-existing research partnerships to facilitate a quick start to data collection. Thus, the three pilot sites represent the three initial languages for the LEVANTE measures (i.e., English, German, and Spanish), a mix of recruitment strategies (i.e., school-based and in-person for Canada and Colombia, family-based and online for Germany), and a mix of high-income (i.e., Canada and Germany) and low/middle-income (i.e., Colombia) countries. Each site is collecting data from at least 300 families, stratified across ages 5 - 12. These pilot data are already being used for assessments of measure reliability and validity.

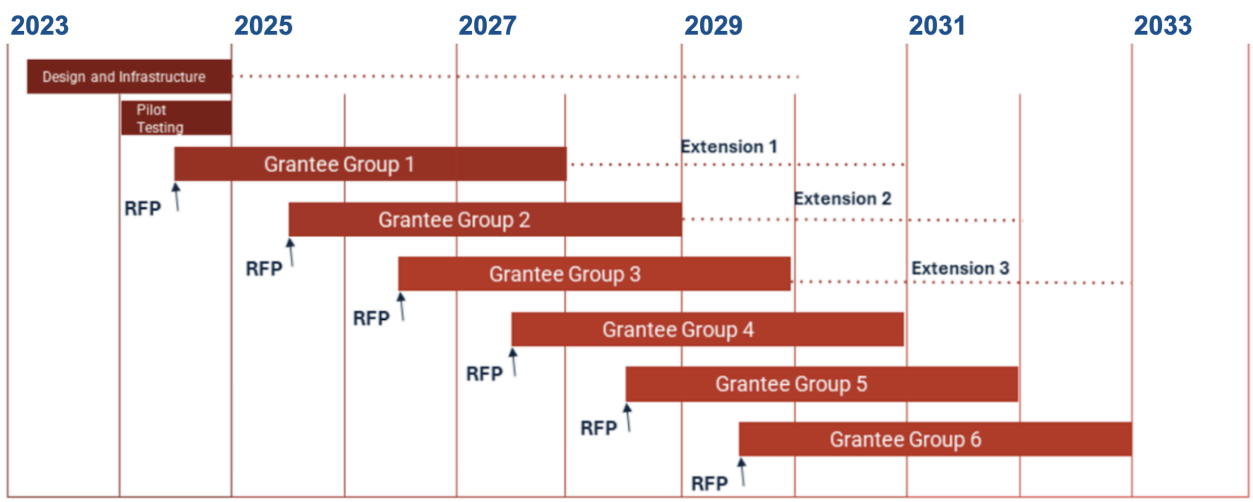

Figure 3. Timeline of anticipated LEVANTE requests for proposals (RFPs) and grant durations.

Building upon the pilot data, there will be a series of calls for proposals, with the first call already published in 2024 and new calls publicized on http://levante-network.org (see Figure 3 for timeline). Within each call for proposals, we anticipate funding approximately 4-8 proposals, leading to an expected total of 20 – 30 culturally, geographically, and linguistically diverse research groups. Review of proposals will prioritize proposed samples that are diverse, representative, from populations that tend to be under-represented in child development research, and relatively large. LEVANTE aims to collect data from a broad set of cultural and linguistic contexts; thus, maximizing the diversity of cultures and languages in the sites is a major factor in site selection. Moreover, representativeness of the sample within the particular context of recruitment (e.g., representativeness of a particular local or national community) is preferred. Finally, larger samples add greater value to the dataset and will be preferred.

Benefits

Below, we list some potential benefits of participating in LEVANTE, at the level of individual researchers and the field of developmental psychology as a whole.

Benefits for individual researchers

The benefits of working on the LEVANTE framework may not be obvious for individual independent researchers. For example, researchers, especially those early in their career, may be evaluated on and vulnerable to metrics that focus primarily on first- and last-author publications. However, we believe that the benefits of participating in large-scale, cross-cultural, longitudinal research outweighs these challenges.

Collaborative projects allow individual researchers to gain experience with ‘best practices’ policies generated by a large community of global researchers. These policies will include recommendations on cross-sectional and longitudinal experimental design, data analysis, and use of collaborative open-science tools. These opportunities to both recommend and learn from best practices policies within a network of global researchers can create additional resources for researchers who may not have the same access in their local institutions.

Large-scale, longitudinal projects yield a vast amount of data. In addition to preregistered analyses that researchers may plan prior to data collection, there are substantial opportunities for the generation of additional research questions and consequent data analyses. Thus, participation in LEVANTE will provide additional publication opportunities from a single dataset.

Finally, LEVANTE’s large-scale, cross-cultural effort will provide researchers around the world with the opportunity to form an international, collaborative intellectual community. Consequently, LEVANTE will provide substantial opportunities for networking, mentorship, and the exchange of research ideas, particularly for researchers from institutions that are currently under-represented in developmental psychology. LEVANTE will provide opportunities for researchers to come together, not only through online exchanges (e.g., virtual meetings), but also through in-person opportunities (e.g., workshops focusing on data collection and analyses).

Benefits for the field as a whole

Psychologists sometimes use different stimuli to measure the same constructs. This practice can lead to difficulty in adjudicating between conflicting results, in other words, determining if conflicting results are caused by failures to replicate an effect or by differences in task and stimulus design. By using identical, or in some cases highly similar (e.g., varying in terms of language), measures across many different ages and contexts, LEVANTE will rule out much of the variability in task and stimulus design, and provide researchers with clearer insight into genuine variability in developmental processes (Cao et al., 2024).

Furthermore, LEVANTE's key goal of prioritizing global participation, both in terms of researchers and research populations, will help to broaden participation of underrepresented geographic, linguistic, cultural, ethnic, and income groups in child development research (e.g., Singh et al., 2023). LEVANTE will further the interdisciplinary field of inquiry into how children grow and develop – in psychology, education, and neuroscience – generating more knowledge about, and consequently helping to empower, the underrepresented communities involved in this research endeavor. By identifying contextual factors that can influence the implementation and scalability of global educational programs, the LEVANTE framework will provide valuable data to inform the creation of education programs that will generalize across global settings (Newbury et al., 2023). Thus, this unique global dataset may help researchers begin to shed light on existing structural inequalities in both the environments and learning outcomes, and may be a valuable step towards reducing these inequalities. Furthermore, LEVANTE could generate empirical evidence to inform global education policies and goals, such as the United Nations' Sustainable Development Goal of Quality Education (i.e., ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all; UNICEF, 2017). These policies may in turn help combat the deeper problems of learning delays (McGregor, 2020) and poor global learning outcomes (Azevedo et al., 2021; Engzell et al., 2021).

Conclusions

LEVANTE is a framework for federated, longitudinal data collection. This framework will support an interrelated set of accelerated, longitudinal studies that use the same data collection platform, core measure set, and data management platform, creating a high-value open dataset for future reuse. We hope through this initiative to shed light on the nature of learning variability during a critical time in development, leading to both theoretical progress in our understanding of development and practical progress in our ability to intervene to improve children's learning outcomes globally.

References

Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, *111*(3), 446–458. https://doi.org/10.1037/edu0000296

Asil, M., & Brown, G. T. L. (2015). Comparing OECD PISA Reading in English to Other Languages: Identifying Potential Sources of Non-Invariance. International Journal of Testing, 16(1), 71–93. https://doi.org/10.1080/15305058.2015.1064431

Azevedo, J.P., Hasan, A., Goldemberg, D., Geven, K., & Iqbal, S.A. (2021). Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer, 36*, 1-40.

Bishop, D. V. (1983). Test for reception of grammar. Published by the author and available from Age and Cognitive Performance Research Centre, University of Manchester, M13 9PL.

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. Normative Data and Psychometric Properties. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(6), 557–564. https://doi.org/10.1097/01.chi.0000159157.57075.c8

Breit, M., Brunner, M., & Preckel, F. (2020). General intelligence and specific cognitive abilities in adolescence: Tests of age differentiation, ability differentiation, and their interaction in two large samples. *Developmental Psychology*, *56*(2), 364–384. https://doi.org/10.1037/dev0000876

Bryan, C.J., Tipton, E. & Yeager, D.S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behavior, 5*, 980–989. https://doi.org/10.1038/s41562-021-01143-3

Call, C. C., Boness, C. L., Cargas, S., & Coakley, K. E. (2024). Measuring Food Security in University Students: A Comparison of the USDA 10-Item and Six-Item Food Security Survey Modules. *Journal of Hunger & Environmental Nutrition*, *0*(0), 1–18. https://doi.org/10.1080/19320248.2024.2310485

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). Early social cognition in three cultural contexts. *Monographs of the Society for Research in Child Development*, *76*(2), vii–viii, 1–142. https://doi.org/10.1111/j.1540-5834.2011.00603.x

Cao, A., Lewis, M., Tsuji, S., Bergmann, C., Cristia, A., & Frank, M. C. (2024, August 23). Estimating age-related change in infants' linguistic and cognitive development using (meta-)meta-analysis. https://doi.org/10.31234/osf.io/qd3cf

Cárdenas, D., Lattimore, F., Steinberg, D., & Reynolds, K. J. (2022). Youth well-being predicts later academic success. *Scientific Reports*, *12*(1), 2134. https://doi.org/10.1038/s41598-022-05780-0

Carstensen, A., Zhang, J., Heyman, G. D., Fu, G., Lee, K., & Walker, C. M. (2019). Context shapes early diversity in abstract thought. *Proceedings of the National Academy of Sciences*, *116*(28), 13891–13896. https://doi.org/10.1073/pnas.1818365116

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464-465.

Davis-Kean, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, *19*(2), 294–304. https://doi.org/10.1037/0893-3200.19.2.294

Dobbs, J., Doctoroff, G.L., Fisher, P.H., & Arnold, D.H. (2006). The association between preschool children's socio-emotional functioning and their mathematical skills. *Journal of Applied Developmental Psychology, 27*, 97-108.

Dornbusch, S. M., Ritter, P. L., Leiderman, P. H., Roberts, D. F., & Fraleigh, M. J. (1987). The Relation of Parenting Style to Adolescent School Performance. *Child Development*, *58*(5), 1244–1257. https://doi.org/10.2307/1130618

Downey, L. A., Lomas, J., Billings, C., Hansen, K., & Stough, C. (2014). Scholastic Success: Fluid Intelligence, Personality, and Emotional Intelligence. *Canadian Journal of School Psychology*, *29*(1), 40–53. https://doi.org/10.1177/0829573513505411

Ellis, B. J., Abrams, L. S., Masten, A. S., Sternberg, R. J., Tottenham, N., & Frankenhuis, W. E. (2022). Hidden talents in harsh environments. *Development and Psychopathology*, *34*(1), 95–113. https://doi.org/10.1017/S0954579420000887

Engzell, P., Frey, A., & Verhagen, M.D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences, 118,* e2022376118.Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Psychometric Properties of the Alabama Parenting Questionnaire. *Journal of Child and Family Studies*, *15*(5), 595–614. https://doi.org/10.1007/s10826-006-9036-y

Essex, M. J., Boyce, W. T., Goldstein, L. H., Armstrong, J. M., Kraemer, H. C., & Kupfer, D. J. (2002). The Confluence of Mental, Physical, Social, and Academic Difficulties in Middle Childhood. II: Developing the MacArthur Health and Behavior Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(5), 588–603. https://doi.org/10.1097/00004583-200205000-00017

Fendinger, N. J., Dietze, P., & Knowles, E. D. (2023). Beyond cognitive deficits: How social class shapes social cognition. *Trends in Cognitive Sciences*, *27*(6), 528–538. https://doi.org/10.1016/j.tics.2023.03.004

Fisher, H. L., Caspi, A., Moffitt, T. E., Wertz, J., Gray, R., Newbury, J., Ambler, A., Zavos, H., Danese, A., Mill, J., Odgers, C. L., Pariante, C., Wong, C. C., & Arseneault, L. (2015). Measuring adolescents' exposure to victimization: The Environmental Risk (E-Risk) Longitudinal Twin Study. *Development and Psychopathology*, *27*(4 Pt 2), 1399–1416. https://doi.org/10.1017/S0954579415000838

Forrest, C.B., Meltzer, L.J., Marcus, C.L., de la Motte, A., Kratchman, A., Buysse, D.J., et al. (2018). Development and Validation of the PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment Item Banks. *Sleep, 41*(6).

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A Collaborative Approach to Infant Research: Promoting Reproducibility, Best Practices, and Theory-Building. *Infancy*, *22*(4), 421–435. https://doi.org/10.1111/infa.12182

Frank, M. C., Braginsky, Mi., Cachia, J., Coles, N. A., Hardwicke, T. E., Hawkins, R. D., Mathur, M. B., & Williams, R. (2024). *Experimentology: An Open Science Approach to Experimental Psychology Methods*. MIT Press. https://doi.org/10.7551/mitpress/14810.001.0001

Frank, M. C., Braginsky, Mi., Yurovsky, D., & Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. MIT Press.

Frankenhuis, W. E., & Nettle, D. (2020). The Strengths of People in Poverty. *Current Directions in Psychological Science*, *29*(1), 16–21. https://doi.org/10.1177/0963721419881154

Frankenhuis, W. E., Young, E. S., & Ellis, B. J. (2020). The Hidden Talents Approach: Theoretical and Methodological Challenges. *Trends in Cognitive Sciences*, *24*(7), 569–581. https://doi.org/10.1016/j.tics.2020.03.007

Frick, P. J., Christian, R. E., & Wootton, J. M. (1999). Age Trends in the Association between Parenting Practices and Conduct Problems. *Behavior Modification*, *23*(1), 106–128. https://doi.org/10.1177/0145445599231005

Garrett-Peters, P. T., Mokrova, I., Vernon-Feagans, L., Willoughby, M., & Pan, Y. (2016). The role of household chaos in understanding relations between early poverty and children's academic achievement. *Early Childhood Research Quarterly*, *37*, 16–25. https://doi.org/10.1016/j.ecresq.2016.02.004

Gilmore, R. O., Adolph, K. E., Millman, D. S., & Gordon, A. (2016). Transforming Education Research Through Open Video Data Sharing. *Advances in Engineering Education*, *5*(2), http://advances.asee.org/wp-content/uploads/vol05/issue02/Papers/AEE-18-Gilmore.pdf.

Golding, J., Pembrey, M., Jones, R., & ALSPAC Study Team. (2001). ALSPAC–the avon longitudinal study of parents and children. *Paediatric and Perinatal Epidemiology*, *15*(1), 74-87.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Green, C. T., Bunge, S. A., Chiongbian, V. B., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematics among children and adolescents. *Journal of Experimental Child Psychology*, *157*, 125–143. https://doi.org/10.1016/j.jecp.2016.12.005

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. American psychologist, 52(10), 1115.

Hammer, D., Melhuish, E., & Howard, S. J. (2018). Antecedents and consequences of social–emotional development: A longitudinal study of academic achievement - ProQuest. *Archives of Scientific Psychology*, *6*. https://www.proquest.com/openview/d89d91a46581c2cc6e9d180686fab77b/1?pq-origsite=gscholar&cbl=2026478

Harden, K. P., Tucker-Drob, E. M., & Tackett, J. L. (2013). The Texas Twin Project. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *16*(1), 385–390. https://doi.org/10.1017/thg.2012.97

Hartung, J., Doebler, P., Schroeders, U., & Wilhelm, O. (2018). Dedifferentiation and differentiation of intelligence in adults across age and years of education. *Intelligence*, *69*, 37–49. https://doi.org/10.1016/j.intell.2018.04.003

Heflin, C., Darolia, R., & Kukla-Acevedo, S. (2022). Exposure to Food Insecurity during Adolescence and Educational Attainment. *Social Problems*, *69*(2), 453–469. https://doi.org/10.1093/socpro/spaa036

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*. https://www2.psych.ubc.ca/~henrich/pdfs/WeirdPeople.pdf

Hughes, C., & Devine, R. T. (2015). Individual Differences in Theory of Mind From Preschool to Adolescence: Achievements and Directions. *Child Development Perspectives*, *9*(3), 149–153. https://doi.org/10.1111/cdep.12124

Jacobs Foundation White Paper (2023). Advancing Research on Learning Variability. Research Agenda. Compiled by Tsang JM, Fetz-Fernandes G and Cubillo A. doi: 10.5281/zenodo.7753301

Janitza, S., Klipker, K., & Hölling, H. (2020). Age-specific norms and validation of the German SDQ parent version based on a nationally representative sample (KiGGS). *European Child & Adolescent Psychiatry*, *29*(2), 123–136. https://doi.org/10.1007/s00787-019-01337-1

Juan-Espinosa, M., Cuevas, L., Escorial, S., & García, L. F. (2006). Testing the indifferentiation hypothesis during childhood, adolescence, and adulthood. *The Journal of Genetic Psychology*, *167*(1), 5–15. https://doi.org/10.3200/GNTP.167.1.5-15

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., et al. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1(4), 443e490.* https://doi.org/10.1177/2515245918810225

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 1–33.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, Article 513. https://doi.org/10.3389/fpsyg.2013.00513

Koita, K., Long, D., Hessler, D., Benson, M., Daley, K., Bucci, M., Thakur, N., & Burke Harris, N. (2018). Development and implementation of a pediatric adverse childhood experiences (ACEs) and other determinants of health questionnaire in the pediatric medical home: A pilot study. *PloS One*, *13*(12), e0208088. https://doi.org/10.1371/journal.pone.0208088

Kühn, S., & Gallinat, J. (2024). Environmental neuroscience unravels the pathway from the physical environment to mental health. *Nature Mental Health, 2,* 263-269.

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, *29*(6), 888–900. https://doi.org/10.1177/0956797617747090

Kulesz, P. A., Roberts, G. J., Francis, D. J., Cirino, P., Walczak, M., & Vaughn, S. (2024). Latent profiles as predictors of response to instruction for students with reading difficulties. *Journal of Educational Psychology, 116*(3), 363–376. https://doi.org/10.1037/edu0000832

Lansford, J. E., & Bornstein, M. H. (2011). Parenting Attributions and Attitudes in Diverse Cultural Contexts: Introduction to the Special Issue. *Parenting*, *11*(2–3), 87–101. https://doi.org/10.1080/15295192.2011.585552

Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., Bacchini, D., Pastorelli, C., Bombi, A.S., Zelli, A., Tapanya, S., Chaudhary, N., Deater-Deckard, K., Manke, B., & Quinn, N. (2005). Physical discipline and children's adjustment: Cultural normativeness as a moderator. *Child Development, 76*, 1234-1246. https://doi.org/10.1111/j.1467-8624.2005.00847.x.

Lansford, J. E., Odgers, C. L., Bradley, R. H., Godwin, J., Copeland, W. E., Rothenberg, W. A., & Dodge, K. A. (2023). The HOME-21: A revised measure of the home environment for the 21st century tested in two independent samples. *Psychological Assessment*, *35*(1), 1–11. https://doi.org/10.1037/pas0001183

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45*(3), 764–781. https://doi.org/10.1037/a0014132

Mabbott, D. J., & Bisanz, J. (2003). Developmental Change and Individual Differences in Children's Multiplication. *Child Development*, *74*(4), 1091–1107. https://doi.org/10.1111/1467-8624.00594

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Lawrence Erlbaum Associates.

Manolitsis, G., Georgiou, G. K., & Tziraki, N. (2013). Examining the effects of home literacy and numeracy environment on early reading and math acquisition. *Early Childhood Research Quarterly*, *28*(4), 692–703. https://doi.org/10.1016/j.ecresq.2013.05.004

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Matheny, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. *Journal of Applied Developmental Psychology*, *16*(3), 429–444. https://doi.org/10.1016/0193-3973(95)90028-4

McGregor, K. K. (2020). How we fail children with developmental language disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981-992.

Micalizzi, L., Brick, L. A., Flom, M., Ganiban, J. M., & Saudino, K. J. (2019). Effects of socioeconomic status and executive function on school readiness across levels of household chaos. *Early Childhood Research Quarterly*, *47*, 331–340. https://doi.org/10.1016/j.ecresq.2019.01.007

Milburn, T.F., Lonigan, C.J., DeFlorio, L., & Klein, A. (2019). Dimensionality of preschoolers' informal mathematical abilities. *Early Childhood Research Quarterly, 47*, 487-495.

Minh, A., Muhajarine, N., Janus, M., Brownell, M., & Guhn, M. (2017). A review of neighborhood effects and early child development: How, where, and for whom, do neighborhoods matter? *Health & Place*, *46*, 155–174. https://doi.org/10.1016/j.healthplace.2017.04.012

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Murphey, D. A. (1992). Constructing the child: Relations between parents' beliefs and child outcomes. *Developmental Review*, *12*(2), 199–232. https://doi.org/10.1016/0273-2297(92)90009-Q

Nag, S., Vagh, S. B., Dulay, K. M., Snowling, M., Donolato, E., & Melby-Lervåg, M. (2024). Home learning environments and children's language and literacy skills: A meta-analytic review of studies conducted in low-and middle-income countries. Psychological Bulletin, 150(2), 132.

Napoli, A. R., & Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: Cross-domain relations of parent–child practices and child outcomes. *Journal of Experimental Child Psychology*, *166*, 581–603. https://doi.org/10.1016/j.jecp.2017.10.002

National Academies of Sciences, Engineering, and Medicine. (2022). *Measuring Sex, Gender Identity, and Sexual Orientation.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26424.

Nelson, K. (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology*, *17*(2), 170–187. https://doi.org/10.1037/0012-1649.17.2.170

Newbury, D. F., Mesa, C., Puglisi, M., Nash, M., Nag, S., Hulme, C., & Snowling, M. J. (2023). Challenges for implementation in diverse settings: Reflections on two randomised controlled trials of educational interventions in South American communities. *Research Papers in Education*, *38*(6), 966-986.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Norbury, C. F., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2016). Younger children experience lower levels of language competence and academic progress in the first year of school: evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(1), 65-73.

Nussenbaum, K., Velez, J. A., Washington, B. T., Hamling, H. E., & Hartley, C. A. (2022). Flexibility in valenced reinforcement learning computations across development. *Child Development*, *93*(5), 1601–1615. https://doi.org/10.1111/cdev.13791

Obradović, J., Sulik, M. J., Yang, Q. T., & Global Executive Function Initiative (GEFI) Network. (2025). Design summary for Same-Different Selection (SDS) Task. Retrieved from https://sparklab.stanford.edu/measures.

Okagaki, L., & Frensch, P. A. (1998). Parenting and Children's School Achievement: A Multiethnic Perspective. *American Educational Research Journal*, *35*(1), 123–144. https://doi.org/10.3102/00028312035001123

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, *10*(1), 96–105. https://doi.org/10.1375/twin.10.1.96

Pettit, G. S., Harrist, A. W., Bates, J. E., & Dodge, K. A. (1991). Family Interaction, Social Cognition and Children's Subsequent Relations with Peers at Kindergarten. *Journal of Social and Personal Relationships*, *8*(3), 383–402. https://doi.org/10.1177/0265407591083005

Pind, J., Gunnarsdóttir, E. K., & Jóhannesson, H. S. (2003). Raven's Standard Progressive Matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences*, *34*(3), 375–386. https://doi.org/10.1016/S0191-8869(02)00058-2

Roche, A. F. (1992). Growth, maturation, and body composition: the Fels Longitudinal Study 1929-1991 (No. 9). Cambridge University Press.

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the Home Learning Environment Across the First 5 Years: Associations With Children's Vocabulary and Literacy Skills at Prekindergarten. *Child Development*, *82*(4), 1058–1075. https://doi.org/10.1111/j.1467-8624.2011.01614.x

Ruzek, E., Jirout, J., Schenke, K., Vitiello, V., Whittaker, J., Pianta, R. (2020). Using self report surveys to measure PreK children's academic orientations: A psychometric evaluation. *Early Childhood Research Quarterly, 50(2)*, 55-66. https://doi.org/10.1016/j.ecresq.2018.10.012

Ryberg, R., Her, S., Temkin, D., Madill, R., Kelley, C., Thompson, J., & Gabriel, A. (2020). Measuring School Climate: Validating the Education Department School Climate Survey in a Sample of Urban Middle and High School Students. *AERA Open*, *6*(3), 2332858420948024. https://doi.org/10.1177/2332858420948024

Saçkes, M., Trundle, K. C., Bell, R. L., & O'Connell, A. A. (2011). The influence of early science experience in kindergarten on children's immediate and later science achievement: Evidence from the early childhood longitudinal study. *Journal of Research in Science Teaching*, *48*(2), 217–235. https://doi.org/10.1002/tea.20395

Schmiedek, F., Lövdén, M., Oertzen, T. von, & Lindenberger, U. (2020). *Within-person structures of daily cognitive performance cannot be inferred from between-person structures of cognitive abilities* (e27576v1). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.27576v1

Schmitt, S. A., Simpson, A. M., & Friend, M. (2011). A longitudinal assessment of the home literacy environment and early language. *Infant and Child Development*, *20*(6), 409–431. https://doi.org/10.1002/icd.733

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis. *Child Development*, *89*(5), 1467–1484. https://doi.org/10.1111/cdev.13068

Seefeldt, C., Denton, K., Galper, A., & Younoszai, T. (1998). Former Head Start Parents' Characteristics, Perceptions of School Climate, and Involvement in Their Children's Education. *The Elementary School Journal*, *98*(4), 339–349. https://doi.org/10.1086/461900

Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, *25*(3), 317–329. https://doi.org/10.1207/s15374424jccp2503_8

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Singh, L., Barokova, M. D., Baumgartner, H. A., Lopera-Perez, D. C., Omane, P. O., Sheskin, M., ... & Frank, M. C. (2024). A unified approach to demographic data collection for research with young children across diverse cultures. Developmental Psychology, 60(2), 211. https://doi.org/10.1037/dev0001623

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy*, *28*(4), 708–737. https://doi.org/10.1111/infa.12545

Smiley, P. A., & Dweck, C. S. (1994). Individual Differences in Achievement Goals among Young Children. *Child Development*, *65*(6), 1723–1743. https://doi.org/10.1111/j.1467-8624.1994.tb00845.x te Brinke, L. W., van de Groep, S., van der Cruijsen, R., & Crone, E. A. (2023). Variability and change in adolescents' prosocial behavior across multiple time scales. *Journal of Research on Adolescence*, *33*(2), 575–590. https://doi.org/10.1111/jora.12827

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research, 49*, 1193-1208.

UNICEF. (2017). *Bridge Fund Quarterly Activity Report.* New York, NY: United Nations Children's Fund.

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., Pérez-Stable, E. J., Riley, W. T., Bloch, M. H., Conway, K., Deeds, B. G., Dowling, G. J., Grant, S., Howlett, K. D., Matochik, J. A., Morgan, G. D., Murray, M. M., Noronha, A., Spong, C. Y., … Weiss, S. R. B. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience*, *32*, 4–7. https://doi.org/10.1016/j.dcn.2017.10.002

Waldfogel, J., Craigie, T.-A., & Brooks-Gunn, J. (2010). Fragile Families and Child Wellbeing. *The Future of Children / Center for the Future of Children, the David and Lucile Packard Foundation*, *20*(2), 87–112.

Ye, M., Hessler, D., Ford, D., Benson, M., Koita, K., Bucci, M., Long, D., Harris, N. B., & Thakur, N. (2023). Pediatric ACEs and related life event screener (PEARLS) latent domains and child health in a safety-net primary care practice. *BMC Pediatrics*, *23*(1), 367. https://doi.org/10.1186/s12887-023-04163-2

Yeatman, J. D., Tang, K. A., Donnelly, P. M., Yablonski, M., Ramamurthy, M., Karipidis, I. I., Caffarra, S., Takada, M. E., Kanopka, K., Ben-Shachar, M., & Domingue, B. W. (2021).

Rapid online assessment of reading ability. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-85907-x

Zettersten, M., Yurovsky, D., Xu, T. L., Uner, S., Tsui, A.S.T., Schneider, R. M., Saleh, A. N., Meylan, S. C., Marchman, V., Mankewitz, J., MacDonald, K., Long, B., Lewis, M., Kachergis, G., Handa, K., deMayo, B., Carstensen, A., Braginsky, M., Boyce, V., Bhatt, N., Bergey, C. A., & Frank, M.C. (2023). Peekbank: An open, large-scale repository for developmental eye-tracking data of children's word recognition. *Behavior Research Methods, 55*, 2485-2500. https://doi.org/10.3758/s13428-022-01906-4

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The Multidimensional Scale of Perceived Social Support. Journal of Personality Assessment, 52(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2

Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., & Berkoff, K. A. (1990). Psychometric Characteristics of the Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, *55*(3–4), 610–617. https://doi.org/10.1080/00223891.1990.9674095