Measurement of Cognition for the National Children's Study

Philip David Zelazo[1]*, Stella F. Lourenco[2], Michael C. Frank[3], Jed T. Elison[1],
Robert K. Heaton[4], Henry M. Wellman[5], Jerry Slotkin[6], Maria Kharitonova[7] and
J. Steven Reznick[8†]

[1] Institute of Child Development, University of Minnesota, Minneapolis, MN, United States, [2] Department of Psychology, Emory University, Atlanta, GA, United States, [3] Department of Psychology, Stanford University, Palo Alto, CA, United States, [4] Department Psychiatry, University of California, San Diego, San Diego, CA, United States, [5] Department of Psychology, University of Michigan, Ann Arbor, MI, United States, [6] Center for Health Research and Translation, University of Delaware, Newark, DE, United States, [7] xSEL Labs, Evanston, IL, United States, [8] Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

The National Children's Study Cognitive Health Domain Team developed detailed plans for assessing cognition longitudinally from infancy to early adulthood. These plans identify high-priority aspects of cognition that can be measured efficiently and effectively, and we believe they can serve as a model for future large-scale longitudinal research. For infancy and toddlerhood, we proposed several paradigms that collectively allowed us to assess six broad cognitive constructs: (1) executive function skills, (2) episodic memory, (3) language, (4) processing speed, (5) spatial and numerical processing, and (6) social cognition. In some cases, different trial sequences within a paradigm allow for the simultaneous assessment of multiple cognitive skills (e.g., executive function skills and processing speed). We define each construct, summarize its significance for understanding developmental outcomes, discuss the feasibility of its assessment throughout development, and present our plan for measuring specific skills at different ages. Given the need for well-validated, direct behavioral measures of cognition that can be used in large-scale longitudinal studies, especially from birth to age 3 years, we also initiated three projects focused on the development of new measures.

Keywords: cognition, measurement, longitudinal, environment, tablet-based

INTRODUCTION

The United States National Children's Study (NCS), authorized by the Children's Health Act of 2000, was designed to examine environmental influences on health and development. From 2009 to 2014, a pilot study (the NCS Vanguard Study) for a planned large-scale longitudinal cohort investigation enrolled over 14,000 participants in 5,000+ families and collected more than 14 million records, including biological and environmental samples. These data are available to qualified researchers via the NCS Archive (NOT-HD-16-005), now located in NICHD's Data and Specimen Hub (DASH). The NCS was terminated by the NIH Director in December 2014, due to concerns about the magnitude, complexity, and long-term feasibility of the originally envisioned project, but prior to this, plans were made to conduct the larger, main study, in which 100,000 children would be assessed longitudinally from before birth to age 21 years. These plans included assessment of a wide range of environmental, biological, and behavioral variables, spanning broad domains of health and human functioning. The current article focuses on the rationale and plans

for assessing cognition throughout development. These plans identify high-priority aspects of cognition that can be measured efficiently and effectively at a wide range of ages, and provide a useful model for future large-scale longitudinal research on the development of cognition across the lifespan. The use of comparable measures across studies, and in different countries, has the potential to add greatly to a richer and more globally representative characterization of cognitive development in relation to both environmental and biological influences, and to important functional outcomes.

Cognition, or thinking, spans a range of information processing skills involved in learning, memory, communication, and problem solving. These skills are essential for healthy adaptation in society, and individual differences in these skills predict important developmental outcomes, including educational achievement, innovation and job success, and parenting and interpersonal relationships (1). Impairments in cognitive skills are markers of numerous disorders emerging in childhood, adolescence, and adulthood (2).

As part of the NCS, the authors of this article were convened to comprise the NCS Cognitive Health Domain Team. Collectively, this team has considerable expertise in the scientific study of a wide range of cognitive skills across the lifespan. The team was charged with proposing and designing assessments of neurocognitive development (from birth to adulthood) that have potential to inform high-priority policy goals, such as the early identification and amelioration of neurodevelopmental disorders, and the amelioration of systemic inequities in developmental outcomes. The proximal aim was to examine influences on key aspects of cognition and the dynamic ways in which cognitive development interacts with processes at other levels of analysis (e.g., environment, brain, genes, epigenetic processes).

Our team proposed to assess six broad cognitive constructs, or types of cognitive skill. These constructs were selected based on an extensive review of research with infants and young children that has shown both that these constructs can be measured directly (behaviorally) in infancy and that they predict important functional outcomes, including educational achievement, and physical and mental health. Assuming 15 min total for the assessment of cognition at each in-person visit and at each age tested, and that not all constructs would be assessed at each visit, we then designed a paradigm-based approach for infants and toddlers, in which participants are assessed using a series of brief (2–3 min) testing paradigms (e.g., visual search for virtual objects that disappear behind a virtual screen; a looking-while-listening assessment of word comprehension). In some cases, several different skills (e.g., cognitive flexibility, inhibitory control, and processing speed) can be assessed simultaneously. That is, different aspects of children's behavioral responses in a single situation can yield measures of different cognitive constructs and skills within those constructs, allowing for increased efficiency in the collection of relevant data.

Paradigms were designed to be presented on a computer tablet, and we planned to record participants' eye movements in response to events during assessments of specific skills. Looking-time measures are especially useful in early infancy, when infants cannot respond verbally and their range of motor responses is limited. For example, at 8 months of age, we planned to employ a computer tablet-based visual delayed-response paradigm, in which objects are (virtually and conspicuously) hidden at one of the locations and then, after a delay, infants are allowed to search for them by looking to a correct location (or at 21 months, by reaching). We planned to vary the duration of the delay and the number of hiding locations to assess how much information can be kept in mind for how long. Performance on switch trials (i.e., when the location of the hidden object is changed after a series of trials at the same location) provides a measure of cognitive flexibility and inhibitory control. Across trials, infants' reaction times to fixate the correct location provide an index of processing speed.

We proposed to assess the following broad cognitive constructs in this paradigm-based way: (1) executive function skills, (2) episodic memory (learning and recall), (3) language, (4) processing speed, (5) spatial and numerical processing, and (6) social cognition. These six types of cognitive skill can be measured straightforwardly starting in infancy, and early individual differences in these skills predict important developmental outcomes, including educational achievement and social functioning. For example, executive function skills measured in childhood have been found to help predict physical health, substance dependence, socioeconomic status, and the likelihood of a criminal conviction at age 32 years, even after controlling for social class of origin and intelligence (IQ) (3). Childhood impairments in executive function skills are markers of numerous disorders emerging in childhood and adolescence. For example, impairments in executive function skills are associated with Attention Deficit/Hyperactivity Disorder, autism spectrum disorder, and Conduct Disorder, among many other conditions (4–6).

In what follows, we define each type of cognitive skill, summarize its significance for understanding functional outcomes across development, discuss the feasibility of its assessment at different ages, and present our plan for measuring specific skills longitudinally from early infancy into adulthood in the context of the planned NCS. We aimed to assess development as continuously as possible from late infancy into adulthood, but specific ages were selected for assessment based on previous research in cognitive development and the need to coordinate our plans with other plans designed for the NCS as a whole

For executive function skills, episodic memory, language, and processing speed, our paradigm-based approach to assessment in infancy and early childhood can be complemented by, and then replaced by, the NIH Toolbox® Cognition Battery, which is presented on a computer tablet (7), and is available here: https://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox/cognition. We also describe three ongoing projects, launched as part of the NCS, that are designed to create new, well-validated, direct behavioral measures of cognition that can be used in future large-scale longitudinal studies, especially from birth to age 3 years. Further information about specific measures can be obtained from the authors, and by consulting the deliverables from the team's planning process (e.g., more detailed plans and examples), which are in

the public domain and can be found at: https://www.nichd.nih.gov/research/supported/NCS.

EXECUTIVE FUNCTION SKILLS

Definition

Executive function skills are the set of attention-regulation skills involved in the deliberate, goal-directed self-regulation of attention and behavior (8). These skills, which depend on the integrity of neural networks involving prefrontal cortex, the anterior cingulate cortex, and other regions [e.g., (9)], include *cognitive flexibility* (i.e., shifting attention flexibly to consider multiple perspectives or to switch between strategies), *inhibitory control* (i.e., deliberately suppressing attention to distractors and refraining from pre-potent impulsive or automatic responding), and *working memory* (i.e., maintaining attention to information in mind and manipulating it during a limited time delay) (10). Executive function skills emerge during the first year of life, develop rapidly between about 2 and 6 years of age, and during the transition to adolescence, and continue to develop into adulthood (11). When considered across the lifespan, executive function skills follow an inverted-U-shaped curve, rising, reaching a peak during the early- or mid-20s, and then falling (12). Executive function deficits in mid- to late- adulthood, such as unwanted intrusions of irrelevant material into one's speech, are signs of neurocognitive decline [e.g., (8, 13)].

Significance

Executive function skills are required for learning and adjustment to school, where children must attend selectively and ignore distractions, keep information in mind, follow rules, suppress impulses to play or aggress, and use imagination creatively and flexibly to solve problems (14). As noted in the Introduction, impairments in executive function skills are markers of numerous disorders emerging in childhood and adolescence (4–6), and individual differences in children's executive function skills are concurrently and predictively related to academic achievement, including math and reading skills [e.g., (15–17)], and to other important functional outcomes including physical health (3). Executive function skills are also strongly positively related to socioeconomic status in childhood [e.g., (18, 19)]. Further research is required to understand the nature of these associations, but there is growing evidence that prefrontally mediated skills, which modulate many other neural functions, may be especially vulnerable to disruption given their dependence on the integrity of these other neural functions, their protracted developmental course, and their sensitivity to the effects of stress [e.g., (4, 6, 20, 21)]. At the same time, however, in children who are at risk for maladaptation, including for low academic achievement and school drop-out, better executive function skills are associated with better outcomes, suggesting that executive function skills are a protective factor and a sign of resilience in the face of adversity [e.g., (22)]. Moreover, the protracted development of executive function skills suggests a relatively long period of neuroplasticity, and evidence indicates that executive function skills can be improved through training in childhood and beyond (23). Executive function skills are therefore seen as potentially critical to addressing socioeconomic, racial, and other "gaps" in opportunity, achievement, and health outcomes, and they are an increasingly popular target of prevention and early intervention programs.

Feasibility

Inhibitory control is essential for *sustained selective attention* (24), and both cognitive flexibility and inhibitory control are needed for *future-oriented planning*, including the predictive *anticipation* of future events (25). These specific skills can be measured straightforwardly in infancy using a simple looking-time procedure presented on a computer tablet (see section Project 1: Developmental Cognitive Profiler (DCP): Design and Creation of a Computer Tablet-Based Assessment of Cognitive Constructs in Infants and Toddlers). For example, it is possible to use eye tracking to measure sustained visual attention to sequences of events that end with a rewarding display, as well as anticipatory looking ahead to future locations when a spatial pattern becomes predictable. At 14 and 24 months, working memory, cognitive flexibility, and inhibitory control can be measured using a tablet-based visual or manual delayed response paradigm [e.g., (26)], which also provides measures of sustained attention and processing speed. For older children, the NIH Toolbox Cognition Battery includes brief (<5 min each) tablet-based measures of working memory (List Sorting Working Memory Test), cognitive flexibility (Dimensional Change Card Sort; DCCS), and inhibitory control (Flanker Inhibitory Control and Attention Test; Flanker) that are available for research purposes, validated, normed, available in Spanish, English, and other languages, and suitable for use across the lifespan, starting as young as 3 or 4 years (27, 28). Downward extensions (DEXTs) of these measures, designed to be continuous with the NIH Toolbox measures, make these tasks more suitable for younger children [(29); see section Project 3: Executive Function Formative Project: Design and Creation of a Developmental Extension (DEXT) of the Executive Function (EF) Measures of the NIH Toolbox Cognition Battery].

Executive Function Measurement Plan

Table 1 shows the measurement plan for this construct. We planned to measure sustained selective attention, cognitive flexibility, and inhibitory control at 8, 14, and 21 months, and predictive (or proactive) attention (i.e., anticipation) at 8 and 14 months, using two paradigms (delayed response and the presentation of predictable patterns, respectively). The NIH Toolbox DCCS would be used to measure cognitive flexibility administered at 3, 4, 5, and 7 years, and then again at 11, 15, and 19 years. The NIH Toolbox Flanker would be used to measure inhibitory control administered at 3, 4, 5, and 9 years, and then again at 13, 17, and 21 years. We planned to measure working memory at 8, 14, and 21 months (using delayed response), and then again at 3, 4, 5, 9, 13, 17, and 21 years using the NIH Toolbox List Sorting Working Memory Test.

EPISODIC MEMORY

Definition

Episodic memory refers to the conscious recollection of past experiences, and is needed for the development of a coherent

TABLE 1 | National Children's Study measurement plan for executive function skills.

Domain	Construct	Measure	0–1 month	8 months	14 months	21 months	3 years	4 years	5 years	7 years	9 years	11 years	13 years	15 years	17 years	19 years	21 years
Cognition	Executive function	Sustained attention		×	×	×											
Cognition	Executive function	Proactive/Predictive attention		×	×												
Cognition	Executive function	Cognitive flexibility		×	×	×	×	×	×	×		×		×		×	
Cognition	Executive function	Inhibitory control		×	×	×	×	×	×	×	×		×		×		×

sense of self, as well as many other important functions (30). This form of memory is also important for learning and future planning. Episodic memory involves encoding, storing, and retrieving information. These processes are interconnected, but they can be described as clusters based on various parameters of memory functionality. From a developmental perspective, one of the earliest emerging clusters is *recognition memory*, defined as differentiating between previously experienced information and novel information. Recognition memory, which provides a measure of encoding and storage in the absence of retrieval demands, continues to develop during infancy and toddlerhood, and appears necessary for the healthy, long-term development of numerous complex cognitive and social skills [e.g., (31)]. Early recognition memory is typically measured using procedures that estimate recognition memory capacity based on the type and amount of information that can be represented, the amount of familiarization needed to form a representation, and the duration of storage. In older children and adults, episodic memory is typically assessed using cued recall or free recall paradigms in addition recognition memory, to distinguish among different processes involved (e.g., storage vs. retrieval). Participants might be asked to reproduce previously experienced events through verbal report or imitative action, for example.

Significance

As noted, episodic memory is essential for a range of functions, and impairments in episodic memory, as seen in cases of anterograde amnesia, are devastating for the individuals involved (32). Individual differences in episodic memory in infancy predict key cognitive outcomes later in life. For example, recognition memory for a series of objects after delays of 1, 3, or 5 min at age 7 months demonstrates the encoding and storage of information, and has been linked to executive function skills assessed at 11 years (33). Recognition memory is affected by brain insults, such as febrile seizures (34), and developmental impairments in recognition memory have important consequences for children's lives. Examples include studies linking poor recognition memory at 7–9 months with perceptual performance at 4.5 years (35), and recognition memory at 6–12 months with adult IQ and academic achievement (31).

Feasibility

Recognition memory can be measured straightforwardly in infancy, during the toddler years, and across development into adulthood. The standard procedure for assessing recognition memory in infants is relatively simple, and is based on research conducted in the early 1970s: two identical visual stimuli are presented briefly then removed. After a brief delay, one of the original stimuli reappears paired with a novel stimulus, and recognition memory is assessed based on visual preference (looking time) for the novel stimulus [(1), for review].

Recognition memory capacity is defined based on parameters, such as the type of information, the duration of initial access to the information, or the duration of storage of the information. Recognition memory capacity typically reaches extremely high levels by 2 years, so we proposed only to assess this construct at

TABLE 2 | National Children's Study measurement plan for memory.

Domain	Construct	Measure	0–1 month	8 months	14 months	21 months	3 years	4 years	5 years	7 years	9 years	11 years	13 years	15 years	17 years	19 years	21 years
Cognition	Memory	Recognition memory		×	×												
Cognition	Memory	Episodic memory				×	×	×		×	×	×				×	
Cognition	Memory	Working memory		×	×	×	×	×	×				×	×	×		×

8 and 14 months. Episodic memory can be assessed reliably and efficiently using the NIH Toolbox Picture Sequence Memory Test from age 3 years and continuing into adulthood (28).

Memory Measurement Plan

Table 2 shows the measurement plan for this construct. We planned to assess recognition memory at 8 and 14 months, and then again at 3, 4, 5, 9, 13, 17, and 21 years. Episodic memory would be measured using an imitation paradigm (in which participants reproduce at demonstrated sequence of actions) at 21 months. The NIH Toolbox Picture Sequence Memory test would be used at 3, 4, 7, 9, 11, and 19 years.

LANGUAGE

Definition

Early language development involves changes in various processes. Early *vocabulary production* (or expressive vocabulary) refers to words that are said and gestures that are formed to communicate. *Vocabulary comprehension* (or receptive vocabulary) refers to words that are understood. Most infants begin to produce some words and gestures late in the first year, and their early production vocabulary is assumed to be a relatively small subset of their comprehension vocabulary [e.g., see (36)]. For example, an infant who produces only 3 or 4 recognizable words or gestures is likely to comprehend the meaning of 30–40 words. *Reading* (including early pre-literacy) refers to a range of skills leading up to and including reading with comprehension.

Significance

Research indicates that children vary widely in their rate of vocabulary growth (37), and that various aspects of early vocabulary growth have long-term implications for subsequent language development and success in important domains (e.g., classrooms). For example, infants' gesture use at age 14 months predicts vocabulary size at 42 months (38), and early vocabulary size predicts outcomes, such as expressive vocabulary at ages 4 and 5 years (39) and language and literacy achievement up to fifth grade (40). Recent research indicates the importance of assessing the developmental trajectory of vocabulary growth. For example, the rate of vocabulary growth between 14 and 46 months (not just vocabulary size at any one age) predicts later vocabulary at 4.5 years, particularly for children from low-socioeconomic status backgrounds (41). Language comprehension and reading are important components of academic achievement that provide a gateway to learning.

Feasibility

A widely used technique for assessing a child's vocabulary is to ask parents whether their child produces and/or comprehends particular words or gestures. This technique has been particularly useful for measuring production vocabulary when parents are asked to select from a categorized list of words that infants might say (42, 43). Parent-reported vocabulary has the weakness, however, of being susceptible to subjective bias (e.g., exaggeration

by parents who over-interpret the linguistic sophistication of their child's verbal output).

Parental assessment of comprehension is more problematic. Comprehension tends to precede production, and parents are unlikely to gain accurate awareness of words that their children comprehend but do not yet utter. At the same time, parents may systematically over- or underestimate their children's presumed comprehension. Thus, while comprehension can be assessed using parent report, there is some advantage to assessing early comprehension directly. One approach is to show the child an array of pictures, ask the child, "Where is the ___?" and code where the child looks. This looking-while-listening technique has been widely used with infants and toddlers to demonstrate developmental differences in comprehension, preservation of individual differences over time, construct validity when compared with parent report, and good test-retest reliability (44).

Vocabulary comprehension can be assessed reliably and efficiently using the NIH Toolbox Picture Vocabulary Test from 3 years of age and continuing into adulthood, while reading can be assessed effectively beginning at about age 7 years with the NIH Toolbox Oral Reading Recognition Test (45).

Language Measurement Plan

Table 3 shows the measurement plan for this construct. We planned to measure vocabulary comprehension at 8, 14, and 21 months, and at 3, 4, 5, 9, 13, 17, and 21 years; vocabulary production at 8, 14, and 21 months, and at 3 years; and reading (recognition/decoding) at 4, 7, 11, 15, and 19 years. A measure of reading comprehension is also recommended at age 17 years, however, because reading comprehension is potentially an important outcome variable, related to academic achievement and other functional outcomes.

PROCESSING SPEED

Definition

Processing speed (PS) refers to the rate at which mental operations are performed. Slower PS is reflected in delayed reaction times to various stimulus arrays and longer times to perform mental activities. It is closely associated with attentional functions, and may reflect efficiency in short-term memory scanning and inspection times for perceptual comparisons (e.g., new vs. old, or same vs. different).

PS shows a well-documented curvilinear relation with age across the lifespan. Similar to the pattern seen with other cognitive processes, including executive function skills and measures of "fluid intelligence," PS shows an increase in efficiency across childhood, peaks in early adulthood, and then on average declines rather sharply through older adulthood. During childhood, PS in different types of tasks improves at a common rate (46–49), supporting the notion of their dependency on more global aspects of brain development. In addition, factor analytic studies provide evidence that various measures of PS are significantly intercorrelated and can be distinguished from other related cognitive constructs, such as response inhibition,

TABLE 3 | NCS measurement plan for language.

Domain	Construct	Measure	0–1 month	8 months	14 months	21 months	3 years	4 years	5 years	7 years	9 years	11 years	13 years	15 years	17 years	19 years	21 years
Cognition	Language	Vocabulary comprehension		X	X	X	X	X	X		X		X				X
Cognition	Language	Vocabulary production		X	X	X	X										
Cognition	Language	Reading (recognition/decoding)						X		X		X		X		X	
Cognition	Language	Reading (comprehension)*													X		

*A measure of reading comprehension, as opposed to reading recognition, is recommended for consideration in late adolescence if resources exist for its development. This measure of reading comprehension can be considered an important outcome variable, likely related to academic achievement.

working memory, and executive function throughout childhood and adolescence (50, 51).

Significance

PS is a fundamental aspect of cognition that might serve as a rate-limiting factor for more complex cognitive skills and for their healthy development (52). Improvements in PS throughout childhood are associated with improved performance on a broad range of cognitive tasks, including executive function skills, reading, memory, arithmetic problem solving, way-finding, and reasoning (53, 54). In some cases, PS accounts for all of the variance in age-related cognitive changes (i.e., a significant association between age and some aspect of cognitive function disappears when PS is partialled out). In addition, longitudinal research shows that measures of PS administered during infancy predict cognitive development in the preschool years (55).

Research has also shown that measures of PS are among the most sensitive cognitive indicators of general cerebral dysfunction (56). For instance, slowed PS has been demonstrated in traumatic brain injury, Multiple Sclerosis, Parkinson's disease, symptomatic human immunodeficiency virus, dementia, and schizophrenia (57). Further, slowed PS has been associated with changes in neurotransmitter activity (e.g., reduced cholinergic function, reduced D2 receptor sites for dopamine, and altered glutamate activity), white matter integrity, glucose metabolism, and nerve conduction velocities (e.g., as measured by evoked potentials, event-related potentials, and electroencephalography) (58). In addition, children with central nervous system impairments, including developmental disability (58), closed head injury (59), and phenylketonuria (60), have slower PS relative to their typically developing peers. Long-term survivors of childhood acute lymphoblastic leukemia who have been treated with radiation (resulting in myelin damage) have slower PS than controls (61, 62). There also is evidence from neuroimaging studies that increased white matter volumes (used as an index of myelination) are associated with faster PS (63). Taken together, results indicate that PS is vulnerable to impairment by a broad range of injuries or diseases involving the brain throughout the lifespan.

Feasibility

Prior research has demonstrated reliable assessment of PS in infants using timed gaze in a continuous familiarization task (64), reaction time in a visual expectation task (65), and a novelty preference score in a visual familiarization paradigm (familiarization time needed before showing a reliable novelty preference) (66). For infants and toddlers, measures of PS can be derived incidentally from performance on tests designed to assess other aspects of cognition, such as executive function, spatial cognition, memory, and social cognition. Looking time responses in the youngest children to various stimuli can be recorded and examined for psychometric suitability as a PS composite. This composite would be conceptually similar to the RT composite created for the NIH Toolbox Cognition Battery (52), and has the added advantage of not requiring time to administer additional, speed-specific items. The latter advantage may be especially important for assessment of the youngest children.

The NIH Toolbox Cognition Battery contains a brief discrimination time (same-different judgments) assessment of PS called the Pattern Comparison Processing Speed Test (52), which has been normed and validated for examinees from ages 7 to 85 years. It has demonstrated strong correlations with results on published tests of PS, and shows adequate test-retest reliability in both children and adults. Moreover, in the large NIH Toolbox norming sample of children aged 3–18 years, the Pattern Comparison Test showed good convergent validity ($r = 0.65$) with a composite reaction time score from two Toolbox tests of executive function (Flanker and DCCS) (52).

Processing Speed Measurement Plan

Table 4 shows the measurement plan for this construct. We planned to measure looking time at 8, 14, and 21 months, and at 3 years; and discrimination time (Pattern Comparison Processing Speed Test) at 3, 4, 5, 7, 9, 13, 17, and 21 years.

SPATIAL AND NUMERICAL PROCESSING

Definition

Spatial cognition, broadly construed, refers to the encoding, manipulation, and use of environmental information (i.e., objects, small-scale configurations, and large-scale, navigable spaces) [e.g., (67–69)]. In human beings, spatial cognition includes behaviors as commonplace as reading a map and as specialized as computing trajectories for the rover Curiosity on Mars. Given the diversity of behaviors supported by spatial cognition, there is growing consensus that it is best understood, particularly over ontogenetic time, when partitioned into well-defined clusters of processes, including: *geometric sensitivity, location coding*, and *mental rotation*. We thus define spatial cognition according to these clusters and briefly describe developmental and individual differences below with reference to each.

1. **Geometric Sensitivity**. Spatial cognition includes the processing of the geometry of individual objects and the geometry of the layout specified by multiple objects (70, 71). Geometric sensitivity encompasses the representation of Euclidean properties, such as length/distance, angle, and sense/direction. Geometric sensitivity supports behaviors, such as shape discrimination, as well as analyses of complex forms and layouts, including comparisons of the relations within and between layouts.
2. **Location Coding**. Spatial cognition also includes the processing of the location of objects, people, and places (72, 73). Location coding involves specifying the location of one or more targets relative to an egocentric and/or allocentric frame of reference. Egocentric reference is specific to the observers themselves (as when an object is specified as being "in front and to my left"), whereas allocentric reference involves landmarks in the external environment and requires use of geometric properties, such as distance and direction (as when an object is specified as being "5 feet southeast of the tower"). To be useful, locations must be

TABLE 4 | NCS measurement plan for processing speed.

Domain	Construct	Measure	0–1 month	8 months	14 months	21 months	3 years	4 years	5 years	7 years	9 years	11 years	13 years	15 years	17 years	19 years	21 years
Cognition	Processing speed	Looking time	X	X	X	X	X										
Cognition	Processing speed	Simple discrimination time				X	X	X	X	X	X	X	X	X	X		X

maintained in memory and updated following changes in viewing perspective.

3. **Mental Rotation**. Spatial cognition, particularly when assessed psychometrically, is most often associated with mental rotation; that is, visualizing and manipulating images of objects, people, or scenes in one's mind, which allows for images to be compared across different perspectives (74, 75). These comparisons can encompass a range of movements, from translations along a single coordinate (e.g., x- or y-axis) to rotations that involve multiple coordinates, including movements in depth.

Given the close conceptual connections between spatial and numerical cognition, as well as accumulating evidence of shared neural mechanisms involving posterior parietal cortex, there is growing consensus among researchers interested in normative and atypical development that these domains are best understood when considered not in isolation, but in relation to each other (76). Measurement skills, and concepts, such as general magnitude representation (77) and the mental number line (78), involve both domains. We thus proposed to assess three aspects of numerical cognition: (1) *number estimation*, (2) *measurement*, and (3) *formal math*.

Numerical cognition ranges from basic abilities, such as estimation (without explicit counting) to more complex skills, such as knowledge of counting routines as well as formal measurement and arithmetic (79). Human and many non-human animals share the capacity to engage in non-verbal number estimation, and this ability is supported by an approximate magnitude system dependent on neural networks involving the intraparietal sulcus, which represents numbers in analog format (80). Although limited in precision, this system has been found to play an important role in the acquisition of symbolic number processes, such as counting and arithmetic (81). The acquisition of measurement skills builds on other basic concepts, such as the mental number line, which changes over the course of childhood (79).

The seeds of spatial and numerical cognition can be measured in infancy. Within the first year, infants discriminate and categorize objects according to shape (82). They also use simple environmental cues, such as the configuration of 2-D forms (83) and distance along a single axis to localize a target (84). Rudimentary mental rotation for 2-D and 3-D objects has been reported in infants between 3 and 9 months (85, 86), though individual differences in performance remain poorly understood (74, 87). In the case of numerical cognition, infants 6–11 months of age not only discriminate arrays of objects that differ in number, but they also represent numerical values according to their ordinal relations (79).

Developmental changes in geometric sensitivity, location coding, mental rotation, and numerical cognition are well-documented within the first 2 years of life, with substantial improvements during the preschool and elementary school years. Studies indicate improvements in location coding, both with respect to the number of locations children can track and the delays they can withstand (88). Children also come to handle more complex allocentric cues, such as distances along multiple,

rather than single, axes (89), and they eventually combine egocentric and allocentric reference frames in optimal (Bayesian) fashion (90). Developmental changes in mental rotation include increases in speed and the complexity of rotation (74). In the case of numerical cognition, significant strides are typically made during the preschool and school years in the acquisition of symbolic number concepts, such as the acquisition of number words, Arabic digits, the use of counting for identifying cardinality (i.e., knowing that the word "five" applies to exactly 5 objects, not 4 or 6) and understanding the successor function (i.e., knowing that the number 5 has only one unique successor, namely, the number 6). Children also tend to acquire a vast range of formal arithmetic concepts, with increases in computational speed and shifts from external to internal mental computations (91, 92).

Individual differences in spatial and numerical cognition are well-documented throughout development, beginning in infancy and continuing into adulthood (93, 94), and children from lower socioeconomic status groups perform worse than children from higher socioeconomic status groups on a variety of measures (95, 96). Socioeconomic status differences can be seen as early as the preschool years and remain stable or increase in magnitude by elementary school age (97, 98).

Significance

Different aspects of spatial cognition, such as geometric sensitivity and mental rotation measured in childhood have been found to be predictive of the likelihood of going on to achieve advanced education credentials and occupations in STEM (Science, Technology, Engineering, and Mathematics) fields, even when controlling for other relevant abilities, such as verbal competence (99). Longitudinal research has found that spatial cognitive abilities measured in elementary school predict creativity and technical innovation later in life, as measured by patents and refereed publications (100). Other studies demonstrate even earlier links between spatial cognition and STEM success (93). Moreover, there is evidence that training mental rotation ability results in improved arithmetic competence in school-age children (101).

Children with delays in learning formal math concepts may be diagnosed with dyscalculia. Children with dyscalculia are generally identified because of difficulties comprehending arithmetic, but their deficits often include more basic abilities, such as non-verbal number estimation (102). In addition to specific learning deficits in math, these children may go on to perform poorly in science and to experience social stigma associated with having a learning disability (103). Recent studies suggest that even basic numerical competence is associated with better occupational prospects and higher income, as well as greater life satisfaction, happiness, and general well-being, even after other cognitive and social factors are controlled (104).

Individuals who struggle with spatial and numerical abilities are at risk for underachieving at school, not securing employment, and lower overall quality of life. These issues not only lead to large personal costs, but also to substantial financial burdens to governments (104). Fortunately, accumulating evidence from training studies suggests that spatial and numerical cognition are not immutable. For example, exposure to puzzles in early childhood and practice with videogames, such as Tetris improve mental rotation performance on standardized tests, and feedback during ordinal tasks (e.g., "Which array contains the larger number of objects?") improves the precision of these judgments, such that smaller differences in numerosity become discriminable (105, 106). Given the reach and impact of spatial and numerical cognition on both our intellectual and social lives, it is increasingly critical that projects, such as the NCS examine the integrity of these constructs from early in development.

Feasibility

The six spatial and numerical constructs we identified for assessment beginning in infancy and across development into adulthood can be measured straightforwardly using standard experimental procedures and a variety of dependent measures [e.g., (82, 84, 93)]. In infancy, these spatial and numerical constructs are assessed using animated displays presented on a computer tablet equipped with a camera for recording looking behaviors, such as anticipatory looking or preferential looking (93). To assess geometry sensitivity, for example, an initial set of trials display a shared property (e.g., shape), and then on subsequent trials, we measure infants' looking behaviors to a novel scenario that requires differentiating between categories (e.g., triangle vs. square). In this example, a character moves behind an occluder on each familiarization trial, with each trial representing a distinct exemplar for the shape category (e.g., three different triangles serve as occluders); during test trials, we measure anticipatory looking toward one of two occluders, one from the same shape category (i.e., novel triangle) and a distractor from a different shape category (e.g., square). A measure of location coding similarly assesses anticipatory looking during test trials, following familiarization to an array of objects (e.g., three containers oriented horizontally) that appears in different positions on the tablet, ensuring an allocentric, not egocentric, target location. Measures of mental rotation and non-verbal number estimation based on preferential looking exist for infants and are predictive of spatial and numerical performance on paper-and-pencil tasks at preschool age (93, 107). In infancy, these tasks are modeled on change detection paradigms, in which two dynamic displays (left and right) vary in the amount of variability. In the mental rotation task, one display includes an object presented from different orientations, whereas the other display also presents the mirrored object. Previous research demonstrates that infants look longer toward the display with the mirrored object, consistent with change detection based on mental rotation. In the non-symbolic number task, one display presents an array of objects that differ in their positions and physical sizes, whereas the other display also presents arrays that differ in number (e.g., alternating displays of 10 and 20 dots). Previous research demonstrates that infants look longer toward the displays with alternating numerosities, when the difference is discriminable (107).

Measures for toddlers, preschoolers, older children, and adults include portable and shortened versions of empirical

assessments used in the existing literature [e.g., (72, 94, 96–98)]. These assessments, presented on a tablet, range from game-like localization tasks for young children to puzzle-like problem solving tasks yielding considerable data (e.g., reaction time data) regarding spatial and numerical processing.

Spatial and Numerical Processing Measurement Plan

Table 5 shows the measurement plan for this construct. We planned to measure geometric sensitivity at 8 and 21 months, and at 3, 5, and 9 years; location coding at 8 and 21 months, and at 3, 4, and 7 years; number estimation at 8 and 14 months, and at 3, 5, and 9 years; mental rotation at 8 and 14 months, and at 3, 5, 9,13, 17, and 21 years; measurement at 3, 4, and 5 years; and formal math at 7, 11, 15, and 19 years.

SOCIAL COGNITION

Definition

Social cognition (i.e., understanding self and other persons) is needed to navigate our complex, human social world. The development of social cognition begins early, as reflected in the *perception and recognition of persons* (e.g., faces, emotions), *social responsiveness* (e.g., joint attention; someone looks at an object and the infant follows that gaze to look at the same object), and *understanding of intentional actions* (i.e., of goal-directed behaviors). Much social cognition centers around a construal of self and others in terms of mental states (e.g., beliefs, desires, intentions, emotions), called *theory of mind;* in colloquial terms, this involves "putting oneself in other peoples' shoes" (appreciating how the other person is likely feeling, interpreting, or reacting to a situation). As adults, we typically construe people as engaging in actions they *believe* will get them what they *desire*: Jill watched TV because she *wanted* entertainment and *thought* a good program was on. Theory-of-mind understandings begin in infancy, and they develop across childhood, as children come to understand more about internal mental states, especially beliefs and desires. A classic assessment for this uses false-belief (FB) tasks. For example: Max places chocolate in the drawer, and when he goes away it is moved to the cupboard. When he returns, where will he look? Success on such tasks shows understanding of crucial distinctions between what's mental and what's real (Max will search in the drawer but the chocolate is really in the cupboard), and a grasp of mental-state subjectivity (I think the chocolate is in the cupboard but Max thinks it's in the drawer). Meta-analyses of hundreds of studies show very young preschoolers fail false-belief tasks (yet succeed on warm-up, comprehension, and control questions) whereas 4-, 5-, and 6-year-olds pass them, and this is true (with modest variations in timing) in communities worldwide (108).

Developmental Sequences

Although they fail false-belief tasks, younger children understand crucial aspects of desires and preferences. This is apparent not only in laboratory tasks (109), but also in children's acquisition and use of *mental state terms* for talking about people. Two- and 3-year-old children talk often about persons' desires via terms like

TABLE 5 | NCS measurement plan spatial and numerical processing.

Domain	Construct	Measure	0–1 month	8 months	14 months	21 months	3 years	4 years	5 years	7 years	9 years	11 years	13 years	15 years	17 years	19 years	21 years
Cognition	Spatial and numerical	Geometric sensitivity		×		×	×		×		×						
Cognition	Spatial and numerical	Location coding		×		×	×	×		×							
Cognition	Spatial and numerical	Number estimation		×	×		×		×		×						
Cognition	Spatial and numerical	Mental rotation		×	×		×		×		×		×		×		×
Cognition	Spatial and numerical	Measurement					×	×	×					×			
Cognition	Spatial and numerical	Formal math								×		×		×		×	

want, need, like, but rarely talk about persons' beliefs. By age 4 and 5 years, children also use terms like *think, know, remember*, and *guess*. A pattern of earlier understanding of desires and later understanding of beliefs is apparent in the conversations of US children, Chinese children (110), Spanish-speaking children (111), and even deaf children speaking sign language (112).

Understanding desires and false beliefs are just two milestones in the protracted development of an extended theory of mind. A standardized Theory of Mind Scale assesses childhood understanding of (1) diverse desires, or "DD" (people can have different desires for the same thing), (2) diverse beliefs, or "DB" (people can have different beliefs about the same situation), (3) knowledge-access, or "KA" (something can be true, but someone might not have access to that information and so be ignorant), (4) false belief, or "FB" (something can be true, but someone might believe something different), and (5) hidden emotion, or "HE" (someone can feel one way but display a different emotion). US preschoolers evidence a clear order of difficulty [as listed above DD>DB>KA>FB>HE; (113)] as do Australian, German, French, Turkish, and Indonesian preschoolers. A very similar (but culturally shaped) sequence characterizes children in China (113) and Iran (114). Deaf children of hearing parents exhibit the same sequence, but greatly delayed, and children with autism are equally dramatically delayed, but moreover exhibit a different sequence (115). This Theory of Mind Scale has been extended to later steps focused on children's understanding of non-literal language (NL), such as sarcastic jokes and shows progressive developments in typically developing children up until about 12 or 13 years, and in delayed individuals (e.g., those with deafness or autism) into adulthood (116).

In summary, we identified a total of four proposed social cognition constructs:

(1) *perception and recognition of persons* (e.g., faces)
(2) *social responsiveness* (e.g., orienting to one's own name, orienting to another's point)
(3) *mental state terms* (including non-literal language)
(4) *theory of mind* (understanding behavior in terms of mental states)

Significance

Social cognition is crucial for the achievement of abilities to communicate, to make friends and affiliations, to cooperate and deceive, to learn from and teach others. Differences in theory of mind understanding, for example, longitudinally predict differences in children's play with their peers, their popularity and social interactions with others including their lying and secret-keeping, and parents' and teachers' ratings of social competence (117–119). Infant social-cognitive understandings longitudinally predict later preschool false-belief understandings even after IQ, language competence, and executive function skills are controlled [e.g., (120)]. Preschool theory of mind predicts children's transition to peer-groups when they later enter school (121), and theory of mind differences correlate with social isolation and difficulties in peer relations in adolescence, for typically developing children and delayed individuals, such as those with autism or deafness (122). Impairments in social

TABLE 6 | NCS measurement plan for social cognition.

Domain	Construct	Measure	0–1 month	8 months	14 months	21 months	3 years	4 years	5 years	7 years	9 years	11 years	13 years	15 years	17 years	19 years	21 years
Cognition	Social	Perception and recognition of persons	×	×													
Cognition	Social	Social responsiveness	×	×	×	×											
Cognition	Social	Theory of mind scale item DD				×	×										
Cognition	Social	Theory of mind scale items DD, DB, KA FB						×									
Cognition	Social	Theory of mind scale items KA, FB, HE							×								
Cognition	Social	ToM scale items FB, HE, sarcasm								×							
Cognition	Social	Mental state terms				×	×			×		×		×		×	
Cognition	Social	Social decision making							×		×		×		×		×

cognition are associated with numerous disorders, including autism (123) and antisocial personality disorder (124).

Feasibility

Standard infant tests exist for the infant constructs, and many of these involve simplified displays (presented on a tablet computer) or parent report measures. For example, eye tracking can be used to assess infants' perception and recognition of persons (125), and for social responsiveness, parents can be asked if the infant "orients/looks when his/her name is called" (this can be tested quickly at an assessment visit as well). Mental state language (for toddlers, preschoolers, and older children) can be assessed via the standardized MacArthur Bates Communicative Development Inventory [CDI; e.g., (37)], a parent-report measure, and tallied from a recording of speech during the assessment visit. The Theory of Mind Scale provides a validated short measure of 6 progressive theory of mind constructs (e.g., DD, DB, KA, FB…) encompassing understandings from 30 months to adolescence, and can be administered via a computer tablet displaying vignettes and via computer adaptive testing (CAT) methods where an abbreviated subset of tasks are presented until consistent successes and failures are found [(116); see section Project 2: Design and Development of a Computerized Version of the Theory of Mind Scale].

Social Cognition Measurement Plan

Table 6 shows the measurement plan for this construct. We planned to measure perception and recognition of persons at 8 and 14 months; social responsiveness at 8, 14, and 21 months; Theory of Mind Scale Item DD at 21 months; Theory of Mind Scale Items DD, DB, KA FB at 3 years; Theory of Mind Scale Items KA, FB, HE at 4 years; Theory of Mind Scale Items FB, HE, NL at 5 years; understanding of mental state terms at 21 months, and 3, 4, 7, 11, 15, and 19 years; and social decision making at 3, 5, 9, 13, 17, and 21 years.

NEXT STEPS

The measurement plans described above were made by a team of domain experts, and they were designed to provide high-priority information about the nature of neurocognitive development from birth to adulthood as part of a large-scale longitudinal study of outcomes of genetic and environmental influences on cognition. These plans can inform the design of other longitudinal studies that include assessment of cognitive function from birth to adulthood.

The NIH Toolbox Cognition Battery works well for many typically developing children starting around ages 3 or 4 years and up, but is not designed to assess all relevant aspects of cognition, such as social cognition, and numerical and spatial processing. Thus, there was a recognized need for new well-validated and standardized direct behavioral measures of cognition that can be used in large-scale longitudinal studies, especially from birth to age 3 years. Three additional and ongoing projects were focused on the development of new behavioral measures for use in such studies, and we encourage the further use, validation, and refinement of these new measures.

FIGURE 1 | A screen shot from an animated sequence designed to measure an aspect of social cognition (social sensitivity) in infants and toddlers.

Deliverables from these projects (e.g., more detailed plans and examples) are in the public domain and can be found at: https://www.nichd.nih.gov/research/supported/NCS.

Project 1: Developmental Cognitive Profiler (DCP): Design and Creation of a Computer Tablet-Based Assessment of Cognitive Constructs in Infants and Toddlers

This Cognitive Health Domain Team project was designed to create standardized assessments of cognition (i.e., executive function, vocabulary, spatial and numerical processing, processing speed, and social cognition) that can be administered via a portable software platform in 15 min or less to infants and toddlers. Animated event sequences (trial sequences) were designed to elicit specific looking and/or reaching patterns indicative of the underlying cognitive constructs. A set of animated event sequences was developed (see **Figure 1**) for use on a computer tablet (see **Figure 2**). The current version of DCP has been refined through an iterative process of testing and tweaking, and is now in validation. Additional potential refinements include:

a. Establish norms for the measures (in English)
b. Develop manuals and training materials
c. Develop multiple forms of the measures
d. Develop Spanish and other language versions and norms
e. Refine the automated scoring algorithms
f. Develop a computer-adaptive version of the DCP measures

Project 2: Design and Development of a Computerized Version of the Theory of Mind Scale

This Cognitive Health Domain Team project involved the creation of a computerized (tablet-based) version of the Theory of Mind Scale (110), a well-validated tabletop test for preschool

FIGURE 2 | A prototype of the Developmental Cognitive Profiler, designed to assess cognition in infants and toddlers by recording gaze in response to animated sequences. Infants are seated on their parent's lap ~60 cm from the tablet. Parents wear infrared glasses.

age children, along with a manual and training materials. The original scope of the project allowed for development of a 6-level version and the validation of the tablet-based measure but was subsequently revised to develop a functional 5-level version that can now be used in try-out research, revised, and subjected to validation and norming.

In the existing Theory of Mind Scale (109, 116), children are presented with short (~2 min) vignettes (acted out using small props) and typically progress through 6 steps in a standard order (i.e., DD>DB>KA>FB>HE>NL). When cross-sectional samples of children are tested, that order reflects the order of difficulty (with many children, including younger ones, passing #1, fewer and only older ones passing #s 4 and 5). In longitudinal testing, individual children typically proceed up the scale in order.

For the NCS, a computerized 5-level version of the standardized Theory of Mind Scale was developed. This scale assesses childhood understanding of (1) diverse desires (DD), (2) diverse beliefs (DB), (3) knowledge-access (KA), (4) false belief (FB), and (5) hidden emotion (HE), which show a clear order of difficulty (as listed) in US, Australian, German, French, Turkish, and Indonesian preschoolers. There are 2 versions of each of the five scenarios, one in a man's voice and one in a woman's voice. These were tested on a tablet computer with 9 children ranging in age from 4 to 9 years. These versions yielded the standard scale pattern described above for the 5-step Scale (i.e., order of difficulty DD>DB>KA>FB>HE). The team also developed a user's manual to direct assessors how to use this version of the Scale, and how to score it (for further information, see https://www.nichd.nih.gov/research/supported/NCS).

These steps alone yield a measure suitable for further validation, and for preliminary use in larger-scale direct behavioral assessments of children's social cognition. Just as for the original scale, children can receive scale scores from 0 to 5 (summing the total correct items for the five items) that are usable in correlational/regression analyses, as well as additional scores useful for more in-depth social cognitive assessments and research (such as research with deaf children or children with autism spectrum disorder, including intervention research where pre-and post-measures include ToM Scale scores).

In a subsequent validation study with a larger sample of children (e.g., N = 80–100), it is recommended that at least a sub-sample of children would receive both the animated tablet version and the original props-and-figurines version of the ToM Scale to directly assess comparability; and a further sub-sample of children would receive alternative versions of the animated version (those with a different mixture of the male and female voices) to establish initial test-retest reliability.

In addition, developing new animations with professional voice recordings for vignette 6 (NL) would allow for broader assessment of ages 11–15 years with this measure. A further useful next step would be adaptation for individualized computer administration (CAT, computer-adaptive testing), so that early responses determine the exact scenarios that are presented (a subset of the 6 possible). Finally, an additional future step should be adapting the animated version for possible large-scale on-line administration.

Project 3: Executive Function Formative Project: Design and Creation of a Developmental Extension (DEXT) of the Executive Function (EF) Measures of the NIH Toolbox Cognition Battery

This project was designed to develop robust and brief measures of executive function (EF) for the NCS that are suitable for a more diverse range of children at the lower end of the ability range, and as young as a mental age of 36 months. This also involved revising the training materials for these measures. This project has yielded two computer tablet-based tasks, the Flanker—Developmental extension (Flanker-DEXT) and the Dimensional Change Card Sort—Developmental extension (DCCS-DEXT), as well as a revised version of the Children's Behavior Questionnaire Very Short Form (CBQ-VSF) (126), which includes a new supplemental scale to serve as a parent report EF measure (CBQ-VSF+EF). The two tablet-based tasks were designed for touchscreen administration as developmental extensions of the comparable NIH Toolbox tasks (for a more diverse sample of children, including those having difficulty with the regular Toolbox versions of Flanker and DCCS). These new measures of EF are fully integrated with the existing Toolbox measures, such that children who fail minimal criteria on the standard Toolbox EF measures are administered the easier DEXT portion of the instrument. Flanker-DEXT contains 30 items and DCCS-DEXT contains 40 items.

The original goal of the DEXT study was to develop, refine, and validate the EF measures, develop materials required to train test administrators, and publish the results of this work. The scope of the project was revised in 2015 such that the measures would be completed and programmed for administration on

an iPad, along with manuals and materials, but no validity or reliability data would be collected.

The developmental extensions of both NIH Toolbox EF measures have been constructed for use on an iPad, and they are accessible through the Apple App Store. These measures are available for use in studies that assess their validity and reliability for a wider range of ages and socio-economic backgrounds (29, 127). Tentative scoring (a sum total of DEXT items answered correctly) has been developed, but will likely be adjusted following the first wave of validation studies.

We recommend a full-scale validation study of the new measures with respect to time burden, usability, reliability, and construct validity in a nationally representative sample of diverse families with preschool-aged children ages 2.5–5.5 years of age. Data analyses should examine correlations of new measures with age, each other, other measures of EF, traditional IQ measures, and criterion measures of school readiness (literacy and numeracy) that have established predictive validity, both overall and with age covaried. Analyses should also examine relations of child scores to parent education level, wealth, income, and other demographic variables.

These validation data would inform future refinements of the measures and scoring protocols, and would provide a foundation for further work to validate the measures in other languages or populations of interest, and to validate them for additional purposes, including early childhood screening and diagnostic purposes. A subsequent step is to create national norms for the new measures using a stratified representative sample.

DISCUSSION AND CONCLUSION

The healthy development of cognitive function, coincident with healthy brain development, is essential for successful adaptation. As such, it is a key dimension of human development to consider both as an outcome of environmental and other influences, and also as a potential protective factor that supports resilience despite risks for adverse outcomes. The National Children's Study Cognitive Health Domain Team designed a plan for assessing the development of cognition longitudinally from infancy through adulthood using a paradigm-based approach in infancy and toddlerhood, and relying largely, but not exclusively, on the NIH Toolbox Cognition Battery starting at around 3–5 years. This plan included direct behavioral assessments of (1) executive function skills, (2) memory, (3) language, (4) processing speed, (5) spatial and numerical processing, and (6) social cognition, administered longitudinally from early infancy into adulthood.

Based on this plan, three research and design projects focused on the development of new direct behavioral measures of cognition that can be used in large-scale longitudinal studies, especially from birth to age 3 years. The projects included: the design and creation of computer-tablet-based and paradigm-based assessment of cognitive constructs in infants and toddlers (the DCP); the design and development of a computerized tablet version of the Theory of Mind Scale; and the design and creation of a Developmental Extension (DEXT) of the executive function

measures of the NIH Toolbox Cognition Battery, together with the creation of a revised version of the Children's Behavior Questionnaire Very Short Form (127), which includes a new scale measuring parent-reported executive function skills. More information about these measures is available at https://www. nichd.nih.gov/research/supported/NCS.

An important feature of the plan for assessing cognition as part of the NCS is the efficiency and practicality with which a wide range of variables can be measured using standardized procedures in which data are recorded automatically. The paradigm-based approach in infancy and toddlerhood relies on well-established laboratory measures of cognition (e.g., visual expectation, delayed response, looking while listening) that have been adapted for use on a computer tablet. With carefully selected trial sequences, several different skills can sometimes be assessed simultaneously, and eye tracking data can be scored using automated scoring algorithms. Starting in early childhood, these same skills can then be measured efficiently and reliably using the NIH Toolbox Cognition Battery, as well as the newly developed computerized version of the Theory of Mind Scale. We believe that these plans can inform the design of future longitudinal studies of cognitive development, and we encourage the use of comparable designs and measures across studies, which has the potential to contribute to a more globally representative characterization of how the development of cognition unfolds.

Other standardized assessments of cognition in infancy and toddlerhood, such as the Bayley Scales of Toddler and Infant Development (128) and the Mullen Scales of Early Learning (129), require high levels of training on the part of the examiners, depend on children's motor skills, and generally take far longer than 15 min. These measures yield relatively global assessments (e.g., cognition, language, motor), may not capture more subtle individual differences, and often fail to predict longer-term outcomes [e.g., (130)].

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data and other deliverables generated for this study can be found at: https://www.nichd.nih.gov/research/supported/NCS.

AUTHOR CONTRIBUTIONS

PZ, SL, MF, JE, RH, HW, JS, MK, and JR drafted the sections of the article and edited the revisions. All authors worked collaboratively as part of the National Children's Study Cognitive Health Domain Team.

FUNDING

REFERENCES

1. Bjorklund DF. Cognitive development: an overview. In: Zelazo PD, editor. *The Oxford Handbook of Developmental Psychology (Vol. 1): Body and Mind*. New York, NY: Oxford University Press (2013). p. 447–76. doi: 10.1093/oxfordhb/9780199958450.013.0016

2. Millan M, Agid Y, Brüne M, Bullmore ET, Carter CS, Clayton NS et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat Rev Drug Discov.* (2012) 11:141–68. doi: 10.1038/nrd3628

3. Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington, et al. A gradient of childhood self-control predicts health, wealth, and public safety. *Proc Natl Acad Sci USA.* (2011) 108:2693–8. doi: 10.1073/pnas.1010076108

4. McLaughlin KA. Future directions in childhood adversity and youth psychopathology. *J Clin Child Adolesc Psychol.* (2016) 45:361–82. doi: 10.1080/15374416.2015.1110823

5. Beauchaine TP, Cicchetti D. Emotion dysregulation and emerging psychopathology: a transdiagnostic, transdisciplinary perspective. *Dev Psychopathol.* (2019) 31:799–804. doi: 10.1017/S0954579419000671

6. Zelazo PD. Executive function and psychopathology: a neurodevelopmental model. *Annu Rev Clin Psychol.* (2020) 16:431–541. doi: 10.1146/annurev-clinpsy-072319-024242

7. Zelazo PD, Bauer PJ. National Institutes of Health Toolbox—Cognition Battery (NIH Toolbox CB): Validation for children between 3 and 15 years. *Monogr Soc Res Child Dev.* (2013) 78:1–172. doi: 10.1111/mono.12039

8. Jacques S, Marcovitch, S. Development of executive function across the life span. In: Overton W, editor. *The Handbook of Life-Span Development*. Hoboken, NJ: John Wiley & Sons, Inc. (2010). p. 431–66. doi: 10.1002/9780470880166.hlsd001013

9. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci.* (2001) 24:167–202. doi: 10.1146/annurev.neuro.24.1.167

10. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn Psychol.* (2000) 41:49–100. doi: 10.1006/cogp.1999.0734

11. Carlson SM, Zelazo PD, Faja S. Executive function. In Zelazo PD, editor. *The Oxford Handbook of Developmental Psychology (Vol. 1): Body and Mind*. New York, NY: Oxford University Press (2013). p. 706–43. doi: 10.1093/oxfordhb/9780199958450.013.0025

12. Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Conway KP, et al. NIH Toolbox Cognition Battery (CB): validation of executive function measures in adults. *J Int Neuropsychol Soc.* (2014) 20:620–9. doi: 10.1017/S1355617714000472

13. von Hippel W. Aging, executive functioning, and social control. *Curr Direct Psychol Sci.* (2007) 16:240–4. doi: 10.1111/j.1467-8721.2007.00512.x

14. Zelazo PD, Blair CB, Willoughby MT. *Executive Function: Implications for Education (NCER 2017–2000)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education (2016). Retrieved from: https://ies.ed.gov/ncer/pubs/20172000/pdf/20172000.pdf

15. Blair C, Razza RP. Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* (2007) 78:647–63. doi: 10.1111/j.1467-8624.2007.01019.x

16. Alloway TP, Gathercole S, Kirkwood, H, Elliott JG. The working memory rating scale: a classroom-based behavioral assessment of working memory. *Learn Individ Differ.* (2009) 19:242–5. doi: 10.1016/j.lindif.2008.10.003

17. Allan NP, Hume LE, Allan DM, Farrington AL, Lonigan, CJ. Relations between inhibitory control and the development of academic skills in preschool and kindergarten: a meta-analysis. *Dev Psychol.* (2014) 50:2368–79. doi: 10.1037/a0037493

18. Merz EC, Wiltshire CA, Noble KG. Socioeconomic inequality and the developing brain: spotlight on language and executive function. *Child Dev Perspect.* (2018) 13:15–20. doi: 10.1111/cdep.12305

19. Hackman DR, Gallop G, Evans W, Farah MJ. Socioeconomic status and executive function: developmental trajectories and mediation. *Dev Sci.* (2015) 18:686–702. doi: 10.1111/desc.12246

20. Shonkoff JP. Protecting brains, not simply stimulating minds. *Science.* (2011) 333:982–3. doi: 10.1126/science.1206014

21. Hair N, Hanson J, Wolfe B, Pollack S. Association of child poverty, brain development, and academic achievement. *JAMA Pediatr.* (2015) 169:822–9. doi: 10.1001/jamapediatrics.2015.1475

22. Masten AS, Herbers JE, Desjardins CD, Cutuli JJ, McCormick CM, Sapienza JK, et al. Executive function skills and school success in young children experiencing homelessness. *Educ Res.* (2012) 41:375–84. doi: 10.3102/0013189X12459883

23. Diamond A, Ling DS. Review of evidence on, and fundamental questions about, efforts to improve executive functions, including working memory. In: Bunting M, Novick J, Dougherty M, Engle RW, editors. *Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience, and Human Development*. New York, NY: Oxford University Press, Oxford Scholarship Online (2019). doi: 10.1093/oso/9780199974467.003.0008

24. Fisher AV. Selective sustained attention: a developmental foundation for cognition. *Curr Opin Psychol.* (2019) 29:248–53. doi: 10.1016/j.copsyc.2019.06.002

25. Haith MM, Hazan C, Goodman GS. Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Dev.* (1988) 59:467–79. doi: 10.2307/1130325

26. Pelphrey KA, Reznick JS, Goldman BD, Sasson N, Morrow J, Donahoe A, et al. Development of visuospatial short-term memory in the second half of the 1st year. *Dev Psychol.* (2004) 40:836–51. doi: 10.1037/0012-1649.40.5.836

27. Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont J, Weintraub S. NIH Toolbox Cognition Battery (CB): measuring executive function and attention. *Monogr Soc Res Child Dev.* (2013) 78:16–33. doi: 10.1111/mono.12032

28. Tulsky DS, Carlozzi NE, Chevalier N, Espy KA, Beaumont JL, Mungas D. Toolbox Cognition Battery (CB): measuring working memory. *Monogr Soc Res Child Dev.* (2013) 78:70–87. doi: 10.1111/mono.12035

29. Distefano R, Fiat AE, Merrick JS, Slotkin J, Zelazo PD, Carlson SM, et al. NIH Toolbox executive function measures with developmental extensions: reliability and validity with preschoolers in emergency housing. *Child Neuropsychol.* (2021). doi: 10.1080/09297049.2021.1888905. [Epub ahead of print].

30. Bauer PJ, Dikmen SS, Heaton RK, Mungas D, Slotkin J, Beaumont JL. NIH Toolbox Cognition Battery (CB): measuring episodic memory. *Monogr Soc Res Child Dev.* (2013) 78:34–48. doi: 10.1111/mono.12033

31. Fagan JF, Holland CR, Wheeler, K. The prediction, from infancy, of adult IQ and achievement. *Intelligence.* (2007) 35:225–31. doi: 10.1016/j.intell.2006.07.007

32. Smith CN, Frascino, JC, Hopkins RO, Squire LR. The nature of anterograde and retrograde memory impairment after damage to the medial temporal lobe. *Neuropsychologia.* (2013) 51:2709–14. doi: 10.1016/j.neuropsychologia.2013.09.015

33. Rose SA, Feldman, JF, Jankowski JJ. Implications of infant cognition for executive functions at age 11. *Psychol Sci.* (2012) 23:1345–55. doi: 10.1177/0956797612444902

34. Martinos MM, Yoong M, Patil S, Chin RFM, Neville BG, Scott RC, et al. Recognition memory is impaired in children after prolonged febrile seizures. *Brain.* (2012) 135:3153–64. doi: 10.1093/brain/aws213

35. Moe V, Smith L. The relation of prenatal substance exposure and infant recognition memory to later cognitive competence. *Infant Behav Dev.* (2003) 26:87–99. doi: 10.1016/S0163-6383(02)00170-4

36. Clark EV, Hecht BF. Comprehension, production, and language acquisition. *Annu Rev Psychol.* (1983) 34:325–49. doi: 10.1146/annurev.ps.34.020183.001545

37. Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, et al. Variability in early communicative development. *Monogr Soc Res Child Dev.* (1994) 59:v-173. doi: 10.2307/1166093

38. Rowe ML, Özçalişkan S, Goldin-Meadow S. Learning words by hand: gesture's role in predicting vocabulary development. *First Lang.* (2008) 28:182–99. doi: 10.1177/0142723707088310

39. Rescorla L, Roberts J, Dahlsgaard, K. Late talkers at 2. *J Speech Lang Hear Res.* (1997) 40:556–66. doi: 10.1044/jslhr.4003.556

40. Lee J. Size matters: Early vocabulary as a predictor of language and literacy competence. *Appl Psycholinguist.* (2010) 32:69–92. doi: 10.1017/S0142716410000299

41. Rowe ML, Raudenbush SW, Goldin-Meadow, S. The pace of vocabulary growth helps predict later vocabulary skill. *Child Dev.* (2012) 83:508–25. doi: 10.1111/j.1467-8624.2011.01710.x

42. Dale PS, Bates E, Reznick JS, Morisset C. The validity of a parent report instrument of child language at twenty months. *J Child Lang.* (1989) 16:239–49. doi: 10.1017/S0305000900010394

43. Reznick JS, Goldsmith L. A multiple form word production checklist for assessing early language. *J Child Lang.* (1989) 16:91–100. doi: 10.1017/S0305000900013453

44. Reznick JS. Visual preference as a test of infant word comprehension. *Appl Psycholinguist.* (1990) 11:145–66. doi: 10.1017/S0142716400008742

45. Gershon RC, Slotkin J, Manly JJ, Blitz DL, Beaumont JL, Schnipke D, et al. NIH Toolbox Cognition Battery (CB): measuring language (vocabulary comprehension and reading decoding). *Monogr Soc Res Child Dev.* (2013) 78:49–69. doi: 10.1111/mono.12034

46. Kail R. Sources of age differences in speed of processing. *Child Dev.* (1986) 57:969–987. doi: 10.2307/1130372

47. Kail R. Developmental functions for speeds of cognitive processes. *J Exp Child Psychol.* (1988) 45:339–64. doi: 10.1016/0022-0965(88)90036-7

48. Kail R. Processing time declines exponentially during childhood and adolescence. *Dev Psychol.* (1991) 27:259–66. doi: 10.1037/0012-1649.27.2.259

49. Sheppard LD, Vernon PA. Intelligence and speed of information-processing: a review of 50 years of research. *Pers Individ Differ.* (2008) 44:535–51. doi: 10.1016/j.paid.2007.09.015

50. Demetriou A, Mouyi A, Spanoudis, G. Modelling the structure and development of G. *Intelligence.* (2008) 36:437–54. doi: 10.1016/j.intell.2007.10.002

51. McAuley T, White D. A latent variables examination of processing speed, response inhibition, and working memory during typical development. *J Exp Child Psychol.* (2011) 108:453–68. doi: 10.1016/j.jecp.2010.08.009

52. Carlozzi NE, Tulsky DS, Kail RV, Beaumont JL. NIH Toolbox Cognition Battery (CB): measuring processing speed. *Monogr Soc Res Child Dev.* (2013) 78:88–102. doi: 10.1111/mono.12036

53. Kail RV. Cognitive development includes global and domain-specific processes. *Merrill Palmer Q.* (2004) 50:445–55. doi: 10.1353/mpq.2004.0031

54. Kail RV. Speed of processing in childhood and adolescence: nature, consequences, and implications for understanding atypical development. In: DeLuca J, Kalmar JH, editors. *Studies on Neuropsychology, Neurology and Cognition. Information Processing Speed in Clinical Populations.* New York, NY: Taylor & Francis (2008). p. 101–23.

55. Rose SA, Feldman JF, Jankowski JJ, Van Rossem R. A cognitive cascade in infancy: pathways from prematurity to later mental development. *Intelligence.* (2008) 36:367–78. doi: 10.1016/j.intell.2007.07.003

56. Hawkins KA. Indicators of brain dysfunction derived from graphic representations of the WAIS-iii/WMS-iii technical manual clinical samples data: a preliminary approach to clinical utility. *Clin Neuropsychol.* (1998) 12:535–51. doi: 10.1076/clin.12.4.535.7236

57. DeLuca J, Kalmar JH. *Information Processing Speed in Clinical Populations.* New York, NY: Taylor & Francis (2008).

58. Nettelbeck T, Wilson C. Speed of information processing and cognition. In: MacLean WE, editor. *Ellis' Handbook of Mental Deficiency, Psychological Theory and Research.* Mahwah, NJ: Erlbaum (1997). p. 245–74.

59. Brookshire B, Levin HS, Song J, Zhang L. Components of executive function in typically developing and head-injured children. *Dev Neuropsychol.* (2004) 25:61–83. doi: 10.1207/s15326942dn2501& 2_5

60. Anderson PJ, Wood SJ, Francis DE, Coleman L, Anderson V, Boneh A. Are neuropsychological impairments in children with early-treated phenylketonuria (PKU) related to white matter abnormalities or elevated phenylalanine levels? *Dev Neuropsychol.* (2007) 32:645–68. doi: 10.1080/87565640701375963

61. Schatz J, Kramer JH, Ablin A, Matthay KK. Processing speed, working memory, and IQ: a developmental model of cognitive deficits following cranial radiation therapy. *Neuropsychology.* (2000) 14:189–200. doi: 10.1037/0894-4105.14.2.189

62. Schatz J, Kramer JH, Ablin A, Matthay KK. Visual attention in long-term survivors of leukemia receiving cranial radiation therapy. *J Int Neuropsychol Soc.* (2004) 10:211–20. doi: 10.1017/S1355617704102075

63. Soria-Pastor S, Gimenez M, Narberhaus A, Falcon C, Botet F, Bargallo N, et al. Patterns of cerebral white matter damage and cognitive impairment in adolescents born very preterm. *Int J Dev Neurosci.* (2008) 26:647–54. doi: 10.1016/j.ijdevneu.2008.08.001

64. Robinson CW, Sloutsky VM. Visual processing speed: effects of auditory input on visual processing. *Dev Sci.* (2007) 10:734–40. doi: 10.1111/j.1467-7687.2007.00627.x

65. Rose SA, Feldman JF, Jankowski JJ. Infant visual recognition memory: independent contributions of speed and attention. *Dev Psychol.* (2003) 39:563–71. doi: 10.1037/0012-1649.39.3.563

66. Domsch H, Lohaus A, Thomas H. Learning and retention in 3- and 6-month-old infants: a comparison of different experimental paradigms. *Eur J Dev Psychol.* (2009) 6:396–407. doi: 10.1080/17405620701269409

67. Hegarty M, Waller DA. Individual differences in spatial abilities. In: Shah P, Miyake A, editors. *The Cambridge Handbook of Visuospatial Thinking.* New York, NY: Cambridge University Press (2005). p. 121–69. doi: 10.1017/CBO9780511610448.005

68. Kritchevsky M. The elementary spatial functions of the brain. In: Stiles-Davis J, Kritchevsky M, Bellugi U, editors. *Spatial Cognition: Brain Bases and Development.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (1988). p. 111–40.

69. Thurstone LL. *Primary Mental Abilities.* Chicago, IL: The University of Chicago Press (1938).

70. Kali Y, Orion, N. Spatial abilities of high-school students in the perception of geologic structures. *J Res Sci Teach.* (1996) 33:369–91. doi: 10.1002/(SICI)1098-273633:4<369::AID-TEA2>3.0.CO;2-Q

71. Spelke E, Lee SA, Izard V. Beyond core knowledge: natural geometry. *Cogn Sci.* (2010) 34:863–84. doi: 10.1111/j.1551-6709.2010.01110.x

72. Vasilyeva M, Lourenco SF. Development of spatial cognition. *Wiley Interdiscipl Rev Cogn Sci.* (2012) 3:349–62. doi: 10.1002/wcs.1171

73. Newcombe NS, Huttenlocher J. Development of spatial cognition. In: Kuhn D, Siegler RS, editors. *Handbook of Child Psychology*, 6th ed. New York, NY: John Wiley & Sons, Inc. (2006). p. 734–76. doi: 10.1002/9780470147658.chpsy0217

74. Lauer JE, Yhang E, Lourenco SF. The development of gender differences in spatial reasoning: a meta-analytic review. *Psychol Bull.* (2019) 145:537–65. doi: 10.1037/bul0000191

75. Shepard RN, Metzler J. Mental rotation of three-dimensional objects. *Science.* (1971) 171:701–3. doi: 10.1126/science.171.3972.701

76. Hubbard EM, Piazza, M, Pinel, P, Dehaene S. Interactions between number and space in parietal cortex. *Nat Rev Neurosci.* (2005) 6:435–48. doi: 10.1038/nrn1684

77. Walsh V. A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn Sci.* (2003) 7:483–8. doi: 10.1016/j.tics.2003.09.002

78. Lourenco SF, Cheung CN, Aulet LS. Is visuospatial reasoning related to early mathematical development? A critical review. In: Henik A, Fias W, editors. *Heterogeneity of Function in Numerical Cognition.* New York, NY: Academic Press (2018). p. 177–210. doi: 10.1016/B978-0-12-811529-9.00010-8

79. Dehaene S. *The Number Sense: How the Mind Creates Mathematics*, revised and updated ed. New York, NY: Oxford University Press (2011).

80. Nieder A, Dehaene S. Representation of number in the brain. *Annu Rev Neurosci.* (2009) 32:185–208. doi: 10.1146/annurev.neuro.051508.135550

81. Siegler RS. Magnitude knowledge: the common core of numerical development. *Dev Sci.* (2016) 19:341–61. doi: 10.1111/desc.12395

82. Quinn PC, Slater AM, Brown E, Hayes RA. Developmental change in form categorization in early infancy. *Br J Dev Psychol.* (2001) 19:207–18. doi: 10.1348/026151001166038

83. Slater A, Mattock, A, Brown E, Bremner JG. Form perception at birth: Cohen and Younger (1984) revisited. *J Exp Child Psychol.* (1991) 51:395–406. doi: 10.1016/0022-0965(91)90084-6

84. Newcombe N, Huttenlocher J, Learmonth A. Infants' coding of location in continuous space. *Infant Behav Dev.* (1999) 22:483–510. doi: 10.1016/S0163-6383(00)00011-4

85. Quinn PC, Liben LS. A sex difference in mental rotation in young infants. *Psychol Sci.* (2008) 19:1067–70. doi: 10.1111/j.1467-9280.2008.02201.x

86. Moore DS, Johnson SP. Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy.* (2010) 16:435–45. doi: 10.1111/j.1532-7078.2010.00058.x

87. Frick A, Möhring W, Newcombe NS. Development of mental transformation abilities. *Trends Cogn Sci.* (2014) 18:536–42. doi: 10.1016/j.tics.2014.05.011

88. Sluzenski J, Newcombe NS, Satlow E. Knowing where things are in the second year of life: implications for hippocampal development. *J Cogn Neurosci.* (2004) 16:1443–51. doi: 10.1162/0898929042304804

89. Vasilyeva M, Huttenlocher J. Early development of scaling ability. *Dev Psychol.* (2004) 40:682–90. doi: 10.1037/0012-1649.40.5.682

90. Nardini M, Jones P, Bedford R, Braddick O. Development of cue integration in human navigation. *Curr Biol.* (2008) 18:689–93. doi: 10.1016/j.cub.2008.04.021

91. Geary DC. Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev Psychol.* (2011) 47:1539–52. doi: 10.1037/a0025510

92. Siegler RS, Robinson M. The development of numerical understandings. *Adv Child Dev Behav.* (1982) 16:241–312. doi: 10.1016/S0065-2407(08)60072-5

93. Lauer JE, Lourenco SF. Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychol Sci.* (2016) 27:1291–8. doi: 10.1177/0956797616655977

94. Halberda J, Ly R, Wilmer JB, Naiman DQ, Germine L. Number sense across the lifespan as revealed by a massive internet-based sample. *Proc Natl Acad Sci USA.* (2012) 109:11116–20. doi: 10.1073/pnas.12001961 09

95. Levine SC, Vasilyeva M, Lourenco SF, Newcombe NS, Huttenlocher J. Socioeconomic status modifies the sex difference in spatial skill. *Psychol Sci.* (2005) 16:841–5. doi: 10.1111/j.1467-9280.2005.01623.x

96. Jordan NC, Huttenlocher J, Levine SC. Differential calculation abilities in young children from middle- and low-income families. *Dev Psychol.* (1992) 28:644–53. doi: 10.1037/0012-1649.28.4.644

97. Siegler RS, Ramani, GB. Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *J Educ Psychol.* (2009) 101:545–60. doi: 10.1037/a0014239

98. Levine SC, Foley A, Lourenco SF, Ehrlich S, Ratliff K. Sex differences in spatial cognition: advancing the conversation. *Wiley Interdiscipl Rev Cogn Sci.* (2016) 7:127–55. doi: 10.1002/wcs.1380

99. Wai J, Lubinski D, Benbow CP. Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J Educ Psychol.* (2009) 101:817–35. doi: 10.1037/a0016127

100. Kell HJ, Lubinski D, Benbow CP, Steiger JH. Creativity and technical innovation: spatial ability's unique role. *Psychol Sci.* (2013) 24:1831–6. doi: 10.1177/0956797613478615

101. Cheung CN, Sung JY, Lourenco SF. Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychol Res.* (2019) 84:2000–17. doi: 10.1007/s00426-019-01202-5

102. Piazza M, Facoetti A, Trussardi AN, Berteletti I, Conte S, Lucangeli D, et al. Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition.* (2010) 116:33–41. doi: 10.1016/j.cognition.2010.03.012

103. Jahoda A, Wilson A, Stalker K, Cairney A. Living with stigma and the self-perceptions of people with mild intellectual disabilities. *J Soc Issues.* (2010) 66:521–34. doi: 10.1111/j.1540-4560.2010.01660.x

104. Butterworth B, Varma S, Laurillard D. Dyscalculia: from brain to education. *Science.* (2011) 332:1049–53. doi: 10.1126/science.1201536

105. Uttal DH, Meadow, NG, Tipton E, Hand LJ, Alden AR, et al. The malleability of spatial skills: a meta-analysis of training studies. *Psychol Bull.* (2013) 139:352–402. doi: 10.1037/a0028446

106. DeWind NK, Brannon EM. Malleability of the approximate number system: effects of feedback and training. *Front Hum Neurosci.* (2012) 6:68. doi: 10.3389/fnhum.2012.00068

107. Starr A, Libertus ME, Brannon, EM. Number sense in infancy predicts mathematical abilities in childhood. *Proc Natl Acad Sci USA.* (2013) 110:18116–20. doi: 10.1073/pnas.1302751110

108. Wellman HM, Cross D, Watson J. Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* (2001) 72:655–84. doi: 10.1111/1467-8624.00304

109. Wellman HM, Liu, D. Scaling of theory-of-mind tasks. *Child Dev.* (2004) 75:523–41. doi: 10.1111/j.1467-8624.2004.00691.x

110. Tardif T, Wellman HM. Acquisition of mental state language in Mandarin- and Cantonese-speaking children. *Dev Psychol.* (2000) 36:25–43. doi: 10.1037/0012-1649.36.1.25

111. Ferres LA. Children's early theory of mind: exploring the development of the concept of desire in monolingual Spanish children. *Dev Sci.* (2003) 6:159–65. doi: 10.1111/1467-7687.00266

112. Anderson D, Reilly J. The MacArthur Communicative Development Inventory: normative data for American sign language. *J Deaf Stud Deaf Educ.* (2002) 7:83–106. doi: 10.1093/deafed/7.2.83

113. Wellman HM, Fang F, Peterson CC. Sequential progressions in a theory-of-mind scale: longitudinal perspectives. *Child Dev.* (2011) 82:780–92. doi: 10.1111/j.1467-8624.2011.01583.x

114. Shahaeian A, Peterson CC, Slaughter V, Wellman HM. Culture and the sequence of steps in theory of mind development. *Dev Psychol.* (2011) 47:1239–47. doi: 10.1037/a0023899

115. Peterson CC, Wellman HM, Liu D. Steps in theory-of-mind development for children with deafness or autism. *Child Dev.* (2005) 76:502–17. doi: 10.1111/j.1467-8624.2005.00859.x

116. Peterson CC, Wellman HM, Slaughter V. The mind behind the message: advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Dev.* (2012) 83:469–85. doi: 10.1111/j.1467-8624.2011.01728.x

117. Eggum ND, Eisenberg N, Kao K, Spinrad TL, Bolnick R, Hofer C, et al. Emotion understanding, theory of mind, and prosocial orientation: relations over time in early childhood. *J Posit Psychol.* (2011) 6:4–16. doi: 10.1080/17439760.2010.536776

118. Razza RA, Blair C. Associations among false-belief understanding, executive function, and social competence: a longitudinal analysis. *J Appl Dev Psychol.* (2009) 30:332–43. doi: 10.1016/j.appdev.2008.12.020

119. Slaughter V, Dennis MJ, Pritchard M. Theory of mind and peer acceptance in preschool children. *Br J Dev Psychol.* (2002) 20:545–64. doi: 10.1348/026151002760390945

120. Wellman HM, Lopez-Duran S, LaBounty J, Hamilton B. Infant attention to intentional action predicts preschool theory of mind. *Dev Psychol.* (2008) 44:618–23. doi: 10.1037/0012-1649.44.2.618

121. Imuta K, Henry JD, Slaughter V, Selcuk B, Ruffman T. Theory of mind and prosocial behavior in childhood: a meta-analytic review. *Dev Psychol.* (2016) 52:1192–205. doi: 10.1037/dev0000140

122. Peterson CC, Slaughter V, Moore C, Wellman HM. Peer social skills and theory of mind in children with autism, deafness or typical development. *Dev Psychol.* (2016) 52:46–57. doi: 10.1037/a0039833

123. Baron-Cohen S, Lombardo MV, Tager-Flusberg H, editors. *Understanding Other Minds: Perspectives From developmental Social Neuroscience.* New York, NY: Oxford University Press (2013). p. 51–71. doi: 10.1093/acprof:oso/9780199692972.001.0001

124. Poletti M, Adenzalo M. Theory of mind in non-autistic psychiatric disorders of childhood and adolescence. *Clin Neuropsychiatry.* (2013) 10:188–95. Available online at: https://www.clinicalneuropsychiatry.org/clinical-neuropsychiatry-volume-10-issue-5-october-2013/

125. Elison JT, Sasson NJ, Turner-Brown LM, Dichter G, Bodfish JW. Age trends in visual exploration of social and nonsocial information in children with autism. *Res Autism Spectrum Disord.* (2012) 6:842–51. doi: 10.1016/j.rasd.2011.11.005

126. Putnam S, Rothbart M. Development of short and very short forms of the Children's Behavior Questionnaire. *J Pers Assess.* (2006) 87:102–12. doi: 10.1207/s15327752jpa8 701_09

127. Kalstabakken AW, Desjardins CD, Anderson JE, Berghuis KJ, Hillyer CK, Seiwert MJ, et al. Executive function measures in early childhood screening: concurrent and predictive validity. *Early Childhood Res Q.* (in press).

128. Bayley N. *Bayley Scales of Infant and Toddler Development*, 3rd ed. San Antonio, TX: The Psychological Corporation (2006).

129. Mullen EM. *Mullen Scales of Early Learning (AGS edition)*. Circle Pines, MN: American Guideline Service Inc. (1995).

130. Brito NH, Fifer WP, Amso D, Barr R, Bell MA, Calkins S, et al. Beyond the Bayley: neurocognitive assessments of development during infancy and toddlerhood. *Dev Neuropsychol.* (2019) 44:220–47. doi: 10.1080/87565641.2018.1564310