Technical Note 54

Young Lives

Measuring Foundational Cognitive Skills in Young Lives using RACER

Jere Behrman, Kristine Briones, Santiago Cueto, Marta Favara, Richard Freund, Annina Hittmeyer, Jennifer Lopez, Nicolás Pazos, Alan Sánchez, Douglas Scott, Margaret Sheridan, and Tassew Woldehanna

Measuring Foundational Cognitive Skills in Young Lives using RACER

Jere Behrman, Kristine Briones, Santiago Cueto, Marta Favara, Richard Freund, Annina Hittmeyer, Jennifer Lopez, Nicolás Pazos, Alan Sánchez, Douglas Scott, Margaret Sheridan, and Tassew Woldehanna

First published by Young Lives in September 2022

Printed on FSC-certified paper from traceable and sustainable sources.

About Young Lives

Young Lives is an international study of childhood poverty and transitions to adulthood, following the lives of 12,000 children in four countries (Ethiopia, India, Peru and Vietnam) since 2001. **www.younglives.org.uk**

Young Lives is funded by UK aid from the UK Government.

Young Lives, Oxford Department of International Development (ODID), University of Oxford, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, UK

Tel: +44 (0)1865 281751 • Email: younglives@qeh.ox.ac.uk

Contents

Author contributions		**4**
		5
		6
1.	**Introduction**	**7**
2.	**The Rapid Assessment of Cognitive and Emotional Regulation (RACER)**	**9**
3.	**The Young Lives RACER sample characteristics**	**14**
4.	**Measuring cognitive abilities with the RACER**	**17**
5.	**Tips for effective use of RACER data in statistical analysis**	**30**
6.	**Database and data dictionary**	**33**
References		**34**
Appendices		**38**

Author contributions

Santiago Cueto, Alan Sánchez, Margaret Sheridan and Tassew Woldehanna contributed to the piloting and data collection phase in 2013. Jere Behrman, Marta Favara and Alan Sánchez conceptualised the document and are the Multiple Principal investigators of the grant funding this technical note. Behrman, Kristine Briones, Favara, Sánchez and Sheridan designed the statistical methodology. Briones, Richard Freund and Nicolás Pazos did the statistical analysis. Briones, Favara, Freund, Sánchez, Sheridan and Douglas Scott contributed to writing the document at different stages. All the co-authors provided comments and agreed that this version is ready for distribution.

Acknowledgements

The authors acknowledge core funding from the UK's Foreign, Commonwealth and Development Office (Grant GB-GOV-1-301108) (for data collection and data analysis), Grand Challenges Canada (Grant 0072-03) (for data collection), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant NICHD R21 HD097576) (for data analysis, including this technical note), and the Old Dart Foundation (Niños del Milenio: Research and Advocacy Grant) (for data analysis).

We would like to express our gratitude to the Young Lives participants for the time dedicated to the study over the years, and to the enumerators and technical teams in the Oxford Department of International Development (UK), Group for the Analysis of Development (GRADE, Peru), Institute of Nutritional Research (IIN, Peru) and the Ethiopian Development Research Institute (EDRI).

We also want to recognise the role that several former and current members/affiliates of the Young Lives team played to facilitate the administration of RACER, and without whom this project would not have been possible. In particular, we would like to thank: Sonya Krutikova for introducing the topic of executive function into Young Lives in 2012 and obtaining initial funding for piloting and fieldwork activities; Jo Boyden for overall support to the international research team; Graham Bray for the management of the project; Javier Escobal and Mary Penny for developing local capacity and support in Peru; Amar Hamoudi for programming RACER and constructing the raw datasets; Monica Lizama for monitoring data collection in Peru; and Mayli Zapata for leading the training of the enumerators in Peru and preparing the enumerators' manual. Hamoudi and Sheridan led the first calculation of RACER outcome variables, in collaboration with Cueto, Krutikova and Sánchez.

Summary

Cognitive skills are a key dimension for child development and for the promotion of learning opportunities, yet substantial challenges remain to measure cognitive skills in large, population-based, samples in low- and middle-income settings. In 2013, the Young Lives study administered RACER, a novel tablet-based test to measure children's foundational cognitive skills (including executive functioning) in Ethiopia and Peru. This technical note has five objectives. First, to describe each of the tasks administered. Second, to describe the protocols used to administer the tasks in both country samples. Third, to describe the cognitive outcomes that are constructed based on the data collected. Fourth, to report on differences in cognitive outcomes by socio-demographic characteristics. Finally, we provide advice for future data users (data to be released in late 2023).

1. Introduction

Cognitive skills refer to the central mental processes involved in the acquisition of knowledge, manipulation of information, and higher-order reasoning. Working together, they gather and process information, and underlie one's memory, learning, attention, decision making, and language abilities (Kiely 2014). Cognitive skills differ in the degree to which they can be intentionally learnt or taught. Skills are often approximated using domain-specific cognitive achievement test scores (for example, test scores in maths, reading comprehension and vocabulary knowledge), which require prior exposure to information, techniques or concepts. For example, literacy skills require prior exposure to the definitions of words and the rules of grammar, and arithmetic skills require prior exposure to symbolic meanings of numerals and operators. These skills are often referred to as crystalised skills. By contrast, there are cognitive skills that are not specific to any single circumscribed topic area, or 'knowledge domain', which do not rely on prior exposure to any set of information or technique. One example of such a skill is *long-term memory*, which is the ability to encode, retain, and retrieve new knowledge.

Prior research has shown that domain-general skills, such as long-term memory, serve as important inputs into domain-specific knowledge and general academic success (Arnon 2019; Blair 2002; Grammer, Coffman, and Ornstein 2013). For example, improving declarative memory makes it easier to learn basic skills like reading and to retain facts. In this technical note, we focus on a subset of four domain-general cognitive skills. Subsets of these skills are often referred to as fluid cognitive skills – they underpin one's ability to reason abstractly and solve problems in novel contexts. However, given their generality, the fact that we examine a slightly larger set of skills than those usually described as 'fluid cognitive capacity', as well as the fact that they serve as inputs into domain-specific skills, we hereafter refer to them as 'foundational' cognitive skills.

A substantial and growing body of research in high-income countries has linked basic cognitive function assessed in laboratory settings to real-world behaviours, demonstrating that, among other functions, individual differences in foundational cognitive skills successfully predict school readiness, knowledge acquisition, and academic achievement (Blair 2002; Blair and Razza 2007; Klingberg 2010; Skinner et al. 1998).

However, major barriers exist that limit the success with which cognitive tests and neuropsychological tasks can be administered in low- and middle-income (LMIC) settings. First, many existing cognitive tasks and neuropsychological tests rely on domain-specific knowledge – including literacy and numeracy – which is markedly variable in LMICs among population groups that face high levels of poverty and low levels of school attainment. Given that we expect capacity in foundational cognitive skills to be independent of this acquired knowledge, testing foundational cognitive skills in ways that do not rely on this acquired knowledge is imperative. Second, many tasks assume shared knowledge among participants; for example, tasks may show respondents pictures of certain animals, weather states, or buildings, which depend on context and are not equally familiar throughout the world. Third, assessments of foundational cognitive skills are commonly operationalised by performance on computer-administrated tasks or short games in laboratory settings (Miyake and Friedman 2012), which can be difficult to administer in LMICs or outside of laboratory settings. They often require specialised software and equipment and take a long time to administer. In contrast, neuropsychological tests are usually administered using ‘paper and pencil’ assessments. In this case, a trained administrator, usually a neuropsychologist with considerable skill and training, engages individuals in a battery of tasks (Ford et al. 2019). The skill required to correctly administer these tasks (training to be a neuropsychologist takes years) makes them difficult to administer with less well-trained enumerators, which may decrease the reliability of measurement in LMIC settings. In addition, a well-constructed neuropsychology battery usually takes several hours to complete. For all the reasons mentioned, data from LMICs on children’s foundational cognitive skills are extremely scarce.

Young Lives is an international study of childhood poverty following 12,000 children in Ethiopia, India (in the states of Andhra Pradesh and Telangana), Peru, and Vietnam since 2002 to provide insights into the changing nature of child poverty at the beginning of the twenty-first century. In 2013, Young Lives administered a novel short test to measure children’s foundational cognitive skills (including executive functioning) in Ethiopia and Peru, a low- and middle-income country, respectively. The tablet-based Rapid Assessment of Cognitive and Emotional Regulation (RACER), developed by Margaret Sheridan and Amar Hamoudi, is a set of cognitive tests that assesses long-term memory, inhibition, working memory, implicit learning, and spatial orienting (Ford et al. 2019; Hamoudi and Sheridan 2015). The focus on two countries – when the project collects data in four countries – was due to budgetary reasons and time constraints.

Five in-person interviews have been completed with Young Lives participants since 2002, with the last in 2016 (Rounds 1–5), and five additional phone survey rounds (Round 6) were administered through the Listening to Young Lives at Work: COVID-19 phone survey in 2020–21, following the COVID-19 pandemic outbreak. The in-person survey rounds consisted of three closely linked components – child, household, and community context surveys (Favara et al. 2021).[1] At the child level, the study has collected rich data on topics such as children’s health (including anthropometrics), education, work, diet diversity, subjective well-being and mental health, aspirations and expectations, tobacco and alcohol consumption, and sexual and reproductive health, among others. Information about the children’s verbal and quantitative abilities was also obtained through the administration of mathematics, reading comprehension, and vocabulary tests. Since 2009 (Round 3), data on the siblings of the Young Lives participants were also collected on selected outcomes (including anthropometrics, vocabulary tests, and psychosocial scales).

One dimension not fully captured by the Young Lives study is related to the acquisition and development of the abilities required to learn and to perform well at school. As discussed above, this is different from achievement test scores, which measure actual achievement based on prior learning. In light of this, in 2013, for the first time, information on foundational cognitive skills was collected via the RACER test. To the best of our knowledge, this is the first time that direct measures of foundational cognitive abilities have been obtained for large samples of children in LMICs. Given the longitudinal nature of the Young Lives sample, these data can be used to understand the determinants of cognitive abilities – for example, the possible role of early-life investments or weather shocks affecting foundation cognitive skills. In parallel, one can also look at the role of these abilities to predict later outcomes, such as test scores, schooling outcomes and risky behaviours.

1 See Young Lives (2017) for more information about the Young Lives research.

This technical note has several objectives. First, to describe each of the tasks administered, including a visualisation of what these tasks looked like on a tablet screen. Second, to describe the protocols used to administer RACER in Ethiopia and Peru. Third, to describe the cognitive outcomes that are constructed based on the data collected. Fourth, to discuss the reliability of the data through comparing the measurements obtained in each country to prior literature and evidence in high-income countries. Specifically, we test for differences in performance by socio-demographic characteristics and expect to find worse performances among younger children and among children from poor backgrounds. Last, we aim to provide advice for future data users, as well as a data dictionary.

2. The Rapid Assessment of Cognitive and Emotional Regulation (RACER)

2.1. Key features

The RACER, a touch screen computer/tablet application designed by Margaret Sheridan and Amar Hamoudi, was introduced in 2013 as part of the Young Lives Round 4 survey in Ethiopia and Peru.[2] The version applied in Young Lives consisted of six different tasks, one of which is presented twice, at the beginning and end of the battery of tasks. These tasks are designed to measure the following skills: long-term memory, inhibition, working memory, implicit learning, and spatial orienting. On average, the total implementation time for these tasks was around 22 minutes among children aged 11-12.[3]

Each task has an identical structure, such that once children get the hang of one task, they should be able to understand the rest easily. In the RACER application, tasks are presented in the form of games. Each game is preceded by a tutorial and a practice period before the children play the game.

Each RACER task is designed with four goals in mind:

a. The task does not require complex language comprehension or literacy or numeracy to perform. This is important in high-poverty settings of LMICs as literacy and numeracy are low, most children speak multiple languages, and the official national language is unlikely to be the native language of many participants. Many existing tasks of cognitive function require a high degree of literacy and are only normed in one language, usually English. In

2 Young Lives selected 20 sites in each country, with oversampling of sites covering poor areas. The sites include both urban and rural areas, representing a range of regions, policy contexts and living conditions that reflect the ethnic, geographical and religious diversities of the countries. For more information on the sampling procedures, see Escobal and Flores (2008) and Outes-León and Sánchez (2008).

3 In some cases, the administration of the RACER was interrupted (e.g., the child had to go to the bathroom, or someone arrived at the house). Calculation of average implementation time excludes cases above 60 minutes so as to calculate the administration time for cases where the administration was not interrupted.

the RACER tasks, instructions are given via short silent videos, reducing the need to read or comprehend verbal or written instructions.

RACER was administered to 97 per cent of the Young Lives index children who were available for interviews in Round 4 and 87 per cent of the siblings sample in Ethiopia and 99 per cent of the index children and the siblings samples in Peru. In order to have a benchmark of how these results compare with the response rates of other instruments, Table 1 also reports the proportion of children who, in Round 4, took the Peabody Picture Vocabulary Test (PPVT), a test that measures receptive vocabulary.[4] The PPVT was administered to 88 per cent of the index children in Ethiopia (and 78 per cent of the siblings) and 99 per cent of the index children (and 98 per cent of the siblings) in Peru. In Ethiopia, the non-response rate of RACER is considerably lower than that of PPVT, while in Peru these rates are similar – and small. It is likely that illiteracy is the reason why close to 1 out of 10 index children in Ethiopia did not take the PPVT. The fact that RACER allows for the measurement of cognitive development of a larger proportion of children in a low-income country setting such as Ethiopia highlights one advantage of RACER over other tests, in which performance is affected by literacy, numeracy, or language.

Table 1. *Response rates*

b. The task does not require a high degree of skill on the part of the fieldworker. This is important in LMIC settings, where educational opportunities are often limited. Many existing tasks of cognitive function require a high degree of enumerator skill; in the United States, neuropsychological tests are often given by high-skilled individuals who have PhDs in clinical psychology, since tests often require judgment calls, multitasking, and knowledge to administer. In RACER, enumerators require little skill to implement the tasks, as instructions are given via short videos and implementation follows simple standardised guidelines.

c. The tasks have been designed using simple abstract shapes and colours that are common across cultures and languages. This is an intentional effort to decrease the impact of culture and language on task performance and comprehension.

d. The tasks are administered on a touch-screen tablet. Most domain-specific cognitive tests, particularly in LMICs, are administered using paper and pencils, which requires laborious data entry and decreases the reliability of measurement. Tests administered on laptops or desktop computers, on the other hand, require children to learn or know how to use a keyboard and mouse before they can participate. The use of tablet computers eliminates these problems, reducing or eliminating the computer knowledge required to play these games and improving measurement.

4 See Cueto et al. (2009) and Cueto and León (2012) for details about the administration of the PPVT to the Young Lives sample.

It appears that efforts in these areas have been successful. RACER has been successfully administered in several countries, beyond Ethiopia and Peru, including Syria, Lebanon, Niger (Chen et al. 2019; Ford et al. 2019; Kim et al. 2020;) and Indonesia. Across these locations, children performed the tasks in ways which indicate that they understand how to perform the tasks, engaged effortfully in game play, and engaged in ways that were similar to how individuals play these tasks in high-income settings (see Section 4). The evidence so far indicates that these tasks can be used for comparable measurement of cognitive functions across diverse settings.

2.2. Task sections

Each RACER task consists of three sections: tutorial, practice, and test.[5]

a. **Tutorial**: Tasks were presented in the form of games. The tutorial section of each game was a simple video that instructs children on how to play the game. The video was silent and showed a hand playing the game or eyes 'looking' at things they need to remember. The video was only played once, so enumerators were instructed to ensure that the child was ready to pay attention when it started.[6]

b. **Practice**: After the tutorial, children had a chance to practice the game just shown to them in the video. The practice section of the game could be repeated up to three times. The idea was that children played the practice test until they met a certain criterion, or the maximum number of repetitions was met. In the RACER, the criterion was that they needed to get 75 per cent correct before they could proceed to the actual test. If they got 75 per cent correct, the application would go straight to the test section of the game. If not, it would go back to the tutorial and then to the practice test for the second time. This cycle of tutorial–practice would continue up to three times or until children achieved the 75 per cent correct criterion. After the third repetition of the practice test, whether or not children achieved the criterion, the game would proceed to the test section. This process was designed to take the responsibility for telling if a child understands the game out of the hands of the experimenter. However, if the experimenter thought the children were confused and believed they could help clarify a rule of the game, it was fine to talk to the children during the practice section. During the test portion, however, the fieldworker was instructed not speak to the children.

c. **Test**: The test section of the game was the main segment and was where the cognitive function of each child was assessed. During this segment, the experimenter should have had minimal interaction with the children. Assistance could have been given if children got distressed, overwhelmed, or frustrated, but this should have been noted by the enumerator.

5 An exception is the final task, since it is a repeat of the first task, which only consists of two sections: practice and test.

6 For a visualisation of the tutorials shown to the participants, see https://www.younglives.org.uk/research-project/cognitive-skills-ethiopia-and-peru or https://www.youtube.com/channel/UCsZkpcrrolOvnzNG2X3Iljg

2.3. Piloting, training and protocol in the Young Lives sample

i RACER administration protocol in the Young Lives sample

For the administration of RACER in the Young Lives sample, the team wanted to make sure that all children understood the instructions. Given that the skills assessments were to be administered to poor children – who were very young in some cases – and many of these children had not used or even seen a touch-screen tablet before, an extra step was added prior to the tutorial section of each task. Before the child viewed the tutorial, the enumerator explained the task to the child using printed pictures that mimicked what the child was about to see on the screen. After the enumerator made sure that the child understood the instructions using these pictures, they then proceeded to the tutorial, practice, and test sections using the tablet (see Section 2.2).

The administration protocol can be summarised as follows:

- **Prior to administration**
  - The enumerator looked for a quiet location to administer the RACER. A truly quiet location was often not available, but enumerators identified a place which was as quiet as possible.
  - The enumerator also made sure that the tablet screen was bright enough, and the tablet was in offline mode.
  - The enumerator and children sat facing each other with their knees close, and the children were asked to put the tablet on their legs.
  - A table was **never** used, even when it was available. This decision was made in order to ensure that administration conditions were as similar as possible for children in both poor and better-off areas.
  - The children were asked to use the same hand – and one only – to press the tablet screen. The choice of hand was according to whether they were left- or right-handed, which was asked at the beginning.
- **During administration**
  - **Enumerator instructions**: For each task, the enumerator explained the instructions to the children using coloured papers that mimicked what the children were about to see on the screen. The enumerator moved their hands around the paper as if it were the tablet screen. The children were then encouraged to do the same.
  - **Tablet instructions (tutorial)**: After the enumerator explained the game, the children were asked to watch the instructions on the tablet, which repeated the enumerator instructions.
  - **Practice**: As in the general instructions, the children were allowed to practice the game three times or until they got the practice test 75 per cent correct. During this period, the enumerator was asked to encourage the children to answer but **not** give the right answers. The enumerator also made sure that, for each trial, the children kept their hand close to the screen (but **not** over the screen) and pressed the screen hard when answering.

– **Test**: The children performed the test with minimal enumerator interaction. If the children got frustrated when answering the tasks, the enumerator was asked to say, ‘don’t worry it’s just a game’ and/or ‘just do it the best you can’. For tasks that involved a quick reaction, the enumerator was also asked to remind the children to return their hand to the same position (close but not over the screen) during the brief period between trials.

ii Training of enumerators and piloting

The enumerators in charge of administering the standard Young Lives questionnaires were also in charge of administering the RACER. In early May 2013, a pre-pilot of RACER was carried out in Peru. This pre-pilot, as in the following ones, did not include any Young Lives respondents. A group of enumerators administered the RACER to six children aged 11-12 (approximately the age of the Young Lives index children in 2013) and six children aged 6-7 (approximately the age of the youngest siblings of the index children in 2013) in a public school in a shantytown in Lima. Based on the pre-pilot results, the programme was slightly adjusted to reduce administration time. Some instructions were shortened and some tasks were simplified. The task pictures printed on paper were also introduced before the tutorials.

In mid-May 2013, 45 enumerators were trained for two days and a pilot with around 25 children took place at Instituto de Investigación Nutricional in Lima City. Local enumerators were also trained on using the RACER application in Ethiopia, and a small pilot took place in the Southern Nations, Nationalities, and People's Region (SNNP). A full pilot of the RACER with the other Round 4 survey instruments took place in one urban, one peri-urban, and two rural sites in the province of Canta, north of Lima City. Approximately 25 children aged 11-12 and their younger siblings aged 7-10 were interviewed using the Young Lives instruments, with each of them completing the RACER tasks at the end of the interview (Figure 1).

Figure 1. *Enumerators administering the RACER tests to children in the province of Canta (region of Lima, Peru)*

3. The Young Lives RACER sample characteristics

The Young Lives sample comprises approximately 2,000 children from the Younger Cohort (born 2001–02) and 1,000 children from the Older Cohort (born 1994–95) in each country. The children were selected from 20 sentinel sites that were defined specifically in each country. The concept of a sentinel site comes from health surveillance studies and is a form of purposive sampling where the site (or 'cluster' in sampling language) is deemed to represent a certain type of population and is expected to show typical trends affecting those people or areas. Semi-random/semi-purposive sampling was done in Ethiopia, India, and Vietnam in choosing the clusters, while in Peru the sampling of clusters was random.[7] In the third survey round in 2009, to understand intra-household differences and dynamics, information on the next youngest sibling of the Younger Cohort children was collected in each country.[8]

The RACER was developed to measure cognitive functions in adults and children aged 6 years and older. Young Lives administered it during the fourth survey round in 2012–13 to the Younger Cohort (index) children who were between 11 and 12 years old and their siblings (Table 1). In Peru, all the siblings interviewed were younger than the index child and 92 per cent of them had the same biological mother and father as the index child. In Ethiopia, on the other hand, only 67 per cent of siblings were younger than the index child, with 94 per cent having the same biological mother and father as the index child.

The RACER was administered to 1,811 index children and 1,216 siblings in Ethiopia (879 younger siblings and 337 older siblings) and 1,878 index children and 756 siblings in Peru. The samples are roughly equally split between females and males. On average, the index children were 12 years old and younger siblings were 8-9 years old in both countries, while the Ethiopian older siblings were, on average, 15-16 years old. Figure 2 shows the age distributions in both countries of the index children and their siblings who performed the RACER.

7 Although sites were selected purposively in Ethiopia, households were selected randomly. More information about the sampling in each country can be found at http://www.younglives.org.uk/content/our-research-methods

8 In Ethiopia, if there was no younger sibling, information was collected on the next older sibling of the index child.

Figure 2. *Age distributions of index children and siblings in Round 4, RACER respondents only*

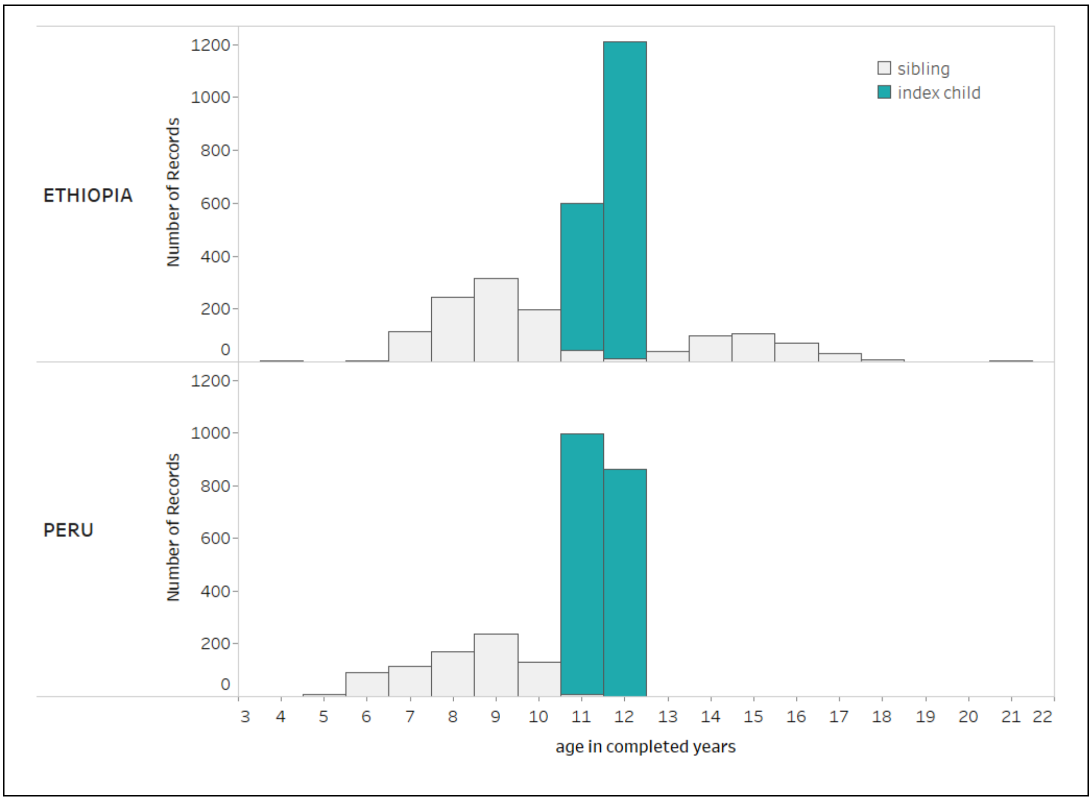


Table 2 outlines the main characteristics of the index children samples.

Table 2. *Round 4 household characteristics of RACER respondents (index children only)*

	Ethiopia			Peru		
	Number of participants	Mean	Std. Dev.	Number of participants	Mean	Std. Dev.
Full sample	**1,812**			**1,878**		
Location						
Urban	670	0.370	0.483	1,378	0.734	0.442
Rural	1,141	0.630	0.483	499	0.266	0.442
Region (Ethiopia only)						
Tigray	380	0.210	0.407			
Amhara	351	0.194	0.395			
Oromia	387	0.214	0.410			
SNNP	425	0.235	0.424			
Addis Ababa	265	0.146	0.354			
Region (Peru only)						
Coast				750	0.400	0.490
Highlands				817	0.435	0.496
Jungle				310	0.165	0.371
Mother's first language (Peru only)						
Spanish				1,318	0.714	0.452
Other				529	0.286	0.452
Mother's highest grade						
No schooling	879	0.502	0.500			
Lower primary	418	0.239	0.426			
Upper primary	280	0.160	0.367			
More than primary	174	0.099	0.299			
Less than primary complete				651	0.359	0.480
Complete primary				528	0.291	0.455
Complete secondary				377	0.208	0.406
Higher education				256	0.141	0.348
Young Lives wealth index						
Quintile 1	407	0.225	0.418	416	0.223	0.416
Quintile 2	370	0.205	0.404	384	0.206	0.404
Quintile 3	362	0.200	0.400	350	0.187	0.390
Quintile 4	326	0.180	0.385	362	0.194	0.395
Quintile 5	343	0.190	0.392	356	0.191	0.393

Notes: The sample includes only index children who completed all RACER tasks. Wealth index values indicate means and standard deviations of the wealth index by quintile. The Young Lives wealth index is a measure of socio-economic status of households within the Young Lives sample that positions the households on a continuous scale of wealth, with higher values reflecting higher household wealth. It is constructed from three indices: housing quality, access to services, and ownership of consumer durables (Briones 2017).

In the following sections, we use information collected for the index children only for simplicity.

4. Measuring cognitive abilities with the RACER

The version of RACER administered in Young Lives comprises six tasks and five games in total (with tasks 1 and 5 grouped as a single game). Each game measures a specific cognitive function: long-term memory (tasks 1 and 5), inhibitory control (task 2), working memory (task 3), and implicit learning (task 4). There is a sixth game (task 6) which is not analysed here (intended to measure spatial orientation).[9] Table 3 presents the tasks and the corresponding cognitive function that they measure.

Table 3. *RACER tasks and cognitive functions*

RACER task #	Cognitive task	Cognitive ability	Definition
Task 1	Paired associate learning task (part 1): 'Memory game 1'	Long-term memory/declarative memory	**Long-term memory/declarative memory:** the ability to encode and retain new knowledge.
Task 2	Simon task: 'Sides game'	Inhibitory control	**Inhibition:** the ability to stop oneself from exhibiting behaviours one does not want to exhibit and is related to one's ability to focus on a single task and supress distractors.
Task 3	Spatial delayed-match-to-sample task: 'Finding the dots'	Working memory	**Working memory:** the ability to hold in mind and manipulate stimuli that are no longer present in the environment.
Task 4	Adapted serial reaction time task: 'Catching chickens/ Chasing dots'	Implicit learning	**Implicit learning:** the ability to recognise and respond to regularities in the environment even when individuals are not aware of these regularities.
Task 5	Paired associate learning task (part 2): 'Memory game 2'	Long-term memory/declarative memory	**Long-term memory/declarative memory:** the ability to encode and retain new knowledge.

The following sections describe each game and the performance measures created for analysis. Each game is composed of multiple rounds, called 'trials'. During the course of each game, participants encounter two separate types of trials – ' baseline 'and 'challenge 'trials – but are not told about any distinction. Baseline trials and challenge trials are identical in terms of general concentration, visual input and motor response, but the baseline trials lack the specific manipulation which requires an individual to employ the foundation cognitive skill under assessment (Ford et al. 2019). A respondent's performance on a baseline trial depends on many factors – including, for instance, familiarity with the use of touchscreen tablets and ability to understand instructions. The game is designed in such a way that challenge trials require the same set of skills and competencies as the baseline trials, plus the specific cognitive skills being assessed. By construction, children are expected to perform worse at

9 Unfortunately, the data from the spatial orienting task were incomplete due to technical difficulties (for example, some responses were simply not recorded on the system).

the challenge trials than the baseline trials.[10] The baseline provides a researcher with information that can be used to sweep out the noise and extract the signal about the relevant foundational cognitive skill.

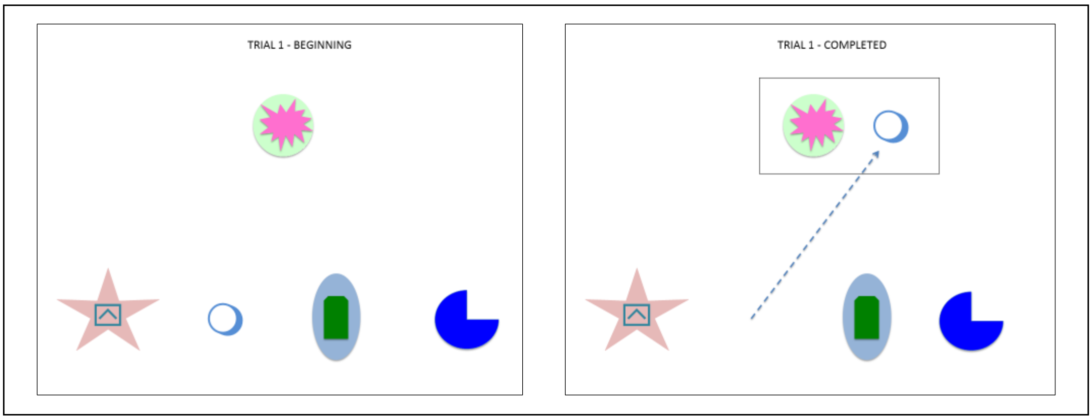
4.1. Memory game: measuring long-term memory

4.1.1. Game description

This game is adapted from the paired associate learning task, which is well-known in cognitive science (Gabrieli 1993; Hannula, Tranel, and Cohen 2006). Children are tasked to determine which shapes go together, in pairs. At first, they will not know which two shapes go together, so they have to guess. However, as the task progresses, and they figure out which shapes go together, they will have an easier time with the game.

The tablet screen shows something similar to that represented in Figure 3. There is a shape shown on the top of the screen and a group of different shapes on the bottom. The first time that respondents see a shape on top of the screen they have no basis on which to identify the correct pair (and thus they will have to choose a pair shape at random). If they touch a lure, the screen does not change, and they will have to choose a new shape. However, if they touch the correct pair at the bottom, the touched shape moves across the screen to join its pair at the top, a box is drawn around the pair, and the two shapes dance in a brief animation sequence. The next trial then begins. In total, there are 12 shapes and six unique pairs (see Figures A1 and A2 for the complete list). Under the memory game, the first appearance of each pair is a baseline trial and subsequent appearances are challenge trials.

Figure 3. *Beginning and completed screens of the memory game*



Respondents see each pair a total of four times. In the first six trials (round 1), each pair is encountered for the first time. In the next six trials (round 2), respondents then cycle through the second encounters with each pair (Table 4). There will then be a (typically) 18–20 minute delay while respondents complete the other tasks of the RACER, with the final task being another two rounds of the memory game where respondents will again match the same pairs as in task 1. In total, the child completes 24 trials in this game.

10 However, in the case of long-term memory the baseline trial is not meant to signal information about a child's abilities. For this task, in the baseline trial the child has to guess the right answer.

Table 4. *Memory game: number of trials per round and task*

Task	Round	Number of trials
1	1	6
	2	6
5	3	6
	4	6
Total number of trials in memory game		24

4.1.2. Performance measure

The rate at which children learn and retain information about these pairs is a measure of long-term memory. Tests of this kind are dependent on the function of the hippocampus in children, adults and animals (Bechara et al. 1995). The hippocampus is a part of the brain which is susceptible to the effects of chronic stress in humans and animals, one reason that it might be expected to see an impact of poverty on the performance of this task (Hanson et al. 2011; Kim and Yoon 1998; McEwen 2001; Shonkoff, Boyce, and McEwen 2009).

Performing better on challenge trials than on baseline trials (the first appearance of each pair) requires long-term memory skill. When comparing baseline to challenge trials, there are two alternative performance measures that may be analysed. The preferred measure is the percentage of trials the child got correct at first touch. A second measure is the total number of touches/choices they made before arriving at the correct paired associate. This second measure is bounded below by 0 and is censored from above by 3, since any respondent who made more than three incorrect guesses must have selected at least one incorrect shape more than once. This truncation and censoring complicate analyses using the second measure. In the Ethiopian and Peruvian samples, 2.3 per cent and 2.5 per cent of the trials involved more than three incorrect guesses, respectively.

Figure 4 shows the preferred performance indicator distributions (the percentages of trials the child got correct at the first touch), comparing the baseline trials (block 1) and challenge trials (blocks 2, 3, and 4). Overall, children performed better in the challenge trails than in the baseline trials. On average, Ethiopian children got 37 per cent of the challenge trails correct at the first touch, compared to 27 per cent of the baseline trials. Peruvian children, on the other hand got, on average, 41 per cent correct at first touch in the challenge trials compared to just 22 per cent correct at first touch in the baseline trials. This observation, that the number of trials correct at first touch increased from the first to last presentation (baseline compared to challenge), is evidence that the task is functioning as expected, as this is the pattern consistently observed in humans and animals in paired associate learning tasks (Eichenbaum 2004).

Figure 4. *Performance indicator distribution: percentages of correct answers at the first touch, Ethiopia and Peru index children*

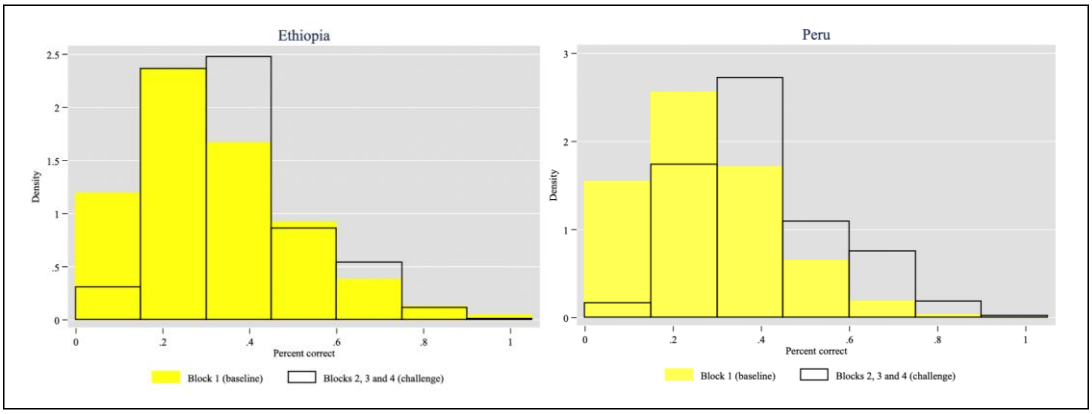


Table B1 reports how the performance measures vary by child and household characteristics in both country samples. As expected, the percentages of correct answers at first touch (in the challenge measure) increase as children get older. In Peru, index children got 41 per cent correct at first touch, whereas their younger siblings only got 37 per cent correct. In Ethiopia, index children got 37 per cent correct at first touch, whereas their younger siblings got 34 per cent correct. This increased capacity to perform a declarative memory across age also replicates findings in high-income countries (Ghetti and Angelini 2008), further demonstrating that the task is functioning as expected in this setting. In Ethiopia (although not in Peru), female index children performed slightly worse than their male counterparts.

Looking at differences according to household characteristics, in both countries, the percentages correct at first touch was higher for children from the top wealth quintile (by 7 and 8 percentage points, respectively) and for children from urban areas (by 5 and 8 percentage points, respectively). The percentages correct at first touch also increased when moving from lower to higher levels of maternal schooling attainment. In high-income countries there is ample evidence that hippocampal function, and by extension paired associate learning capacity, improves with increasing wealth (Farah et al. 2006; Hackman and Farah 2009).

4.2. Sides game: measuring inhibitory control

4.2.1. Game description

This game is based on the ‘Simon task’, a game developed by Simon and Rudell (1967). In each trial, the child is presented with a dot on either the left or the right side of the screen. The dot is either solid and yellow or striped and pink. If the dot is solid and yellow, the child should simply touch the dot as close as possible to its centre. If the dot is striped and pink, however, the child should touch the opposite side of the screen, as close as possible to where the centre of the dot’s mirror image would be. The dot disappears after 2.5 seconds or as soon as the child touches the screen (whichever comes first). The screen shows something similar to that presented in Figure 5. In total, 30 yellow same-side trials and 30 pink opposite-side trials are presented (Table 5).

Figure 5. *Same-side and opposite-side screens of the sides game*

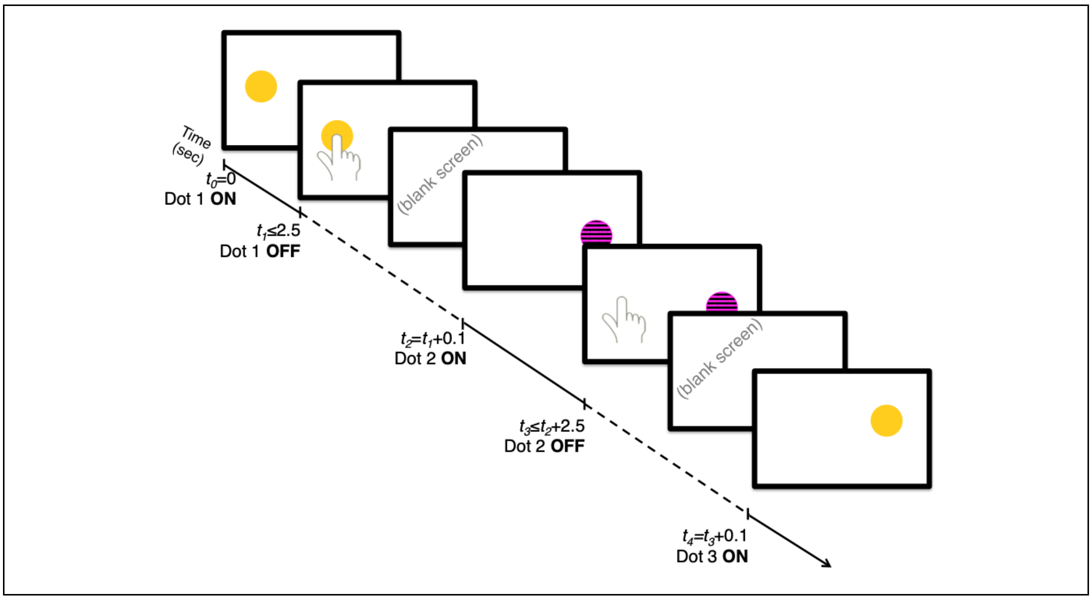


Table 5. *Number of trials in the sides game*

Type	Total number of trials
Same side	30
Opposite side	30
Total	60

4.2.2. Performance measure

Inhibition is the ability to withhold responses (behavioural/attentional) to stimuli in the environment that are likely to capture attention and elicit a response. Conceptually, this cognitive function underlies one's ability to stay on task despite the existence of distracting stimuli, and to stop oneself from exhibiting behaviours one does not want to do. Assessed skill of inhibitory control has been observed to correlate with impulse control in and outside of the classroom (Barkley, Grodzinsky, and DePaul 1992).

The sides game is administered to assess inhibitory control. The children are shown two kinds of dots: yellow dots, which they should press the centre of, and pink striped dots, where they should press on the opposite side of the screen. Because pressing an object (as required in the same-side trials) is a more common reaction than pressing away from an object, correctly performing the opposite-side trials requires children to supress the more prepotent behavioural response in the service of task goals. This suppression is a measure of inhibitory control.

Performance measures of this game include continuous measures of response/reaction time (in seconds) and accuracy (Euclidean distance of touch from centre of dot, in pixels) on same-side (baseline) trials, relative to opposite-side (challenge) trials. On the opposite-side trials, respondents must inhibit their impulse to touch the stimulus as soon as it appears, and instead redirect their movement toward the less salient side of the screen where there is no stimulus drawing their attention.

Figure 6 shows the response time performance distributions of baseline (same-side) and challenge (opposite-side) trials. As expected, children perform worse on challenge trials in both countries. On average, children in the Ethiopian sample responded to same-side trials in 1.13 seconds, and to opposite side trials in 1.18 seconds. Peruvian children responded to same side trials in 1.08 seconds, and to opposite side trials in 1.11 seconds.

Figure 6. *Performance indicator distributions: response time (in seconds), Ethiopia and Peru index children*

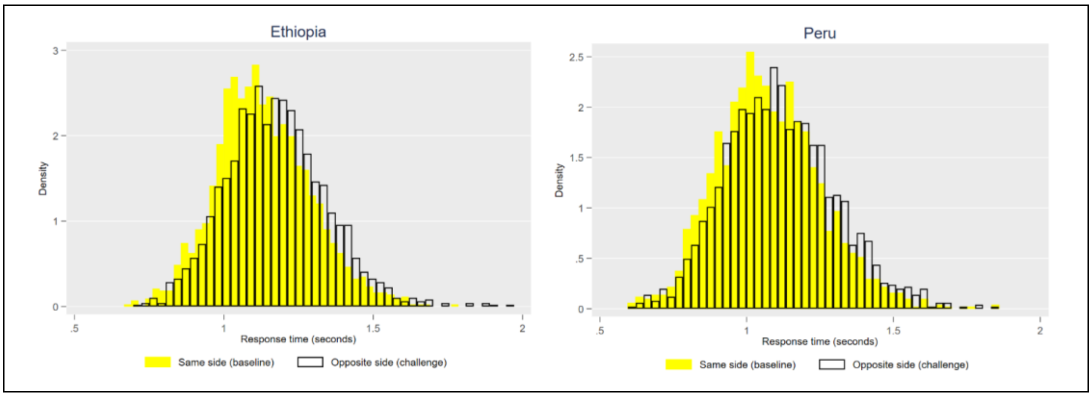


Table B2 reports the differences by child and household characteristics. As expected, response times were lower for older children. In the Peruvian sample, index children had an average response time of 1.11 seconds, compared to 1.28 among their younger siblings. Similarly, in the Ethiopian sample, index children had an average response time of 1.18 seconds, compared to 1.27 among their younger siblings. In Peru (but not Ethiopia), the average response time also increased when the (index child) respondent was female. In terms of household characteristics, as expected, response times were lower for children from the top wealth quintile when compared to the lowest quintile (time reduced by 0.15 and 0.07 seconds, respectively), and among children from urban areas (by 0.72 and 0.05 seconds, respectively). Response time also decreased with maternal schooling attainment; moving from the lowest to the highest category of maternal schooling was associated with decreased response times by 0.13 and 0.11 seconds in Peru and Ethiopia, respectively. Again, these findings replicate what is observed in high-income countries using inhibitory control tasks (Noble, McCandliss, and Farah 2007; Sheridan et al. 2014;).

Figure 7 shows the distributions of the logarithms of the Euclidean distances from touch to correct location for baseline (same-side) and challenge (opposite-side) trials. As can be seen, children from both countries were considerably less accurate on opposite-side trials than same-side trials. Children in the Ethiopian sample were 36 per cent less accurate on opposite-side trials, while Peruvian children were 42 per cent less accurate on opposite-side trials.

Figure 7. *Performance indicator distributions: Euclidean distances from touch to correct location (log scale), Ethiopia and Peru index children*

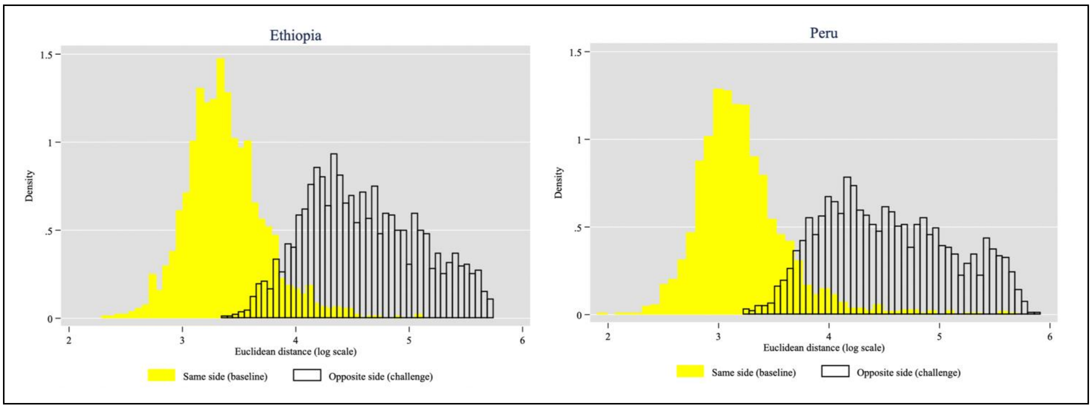


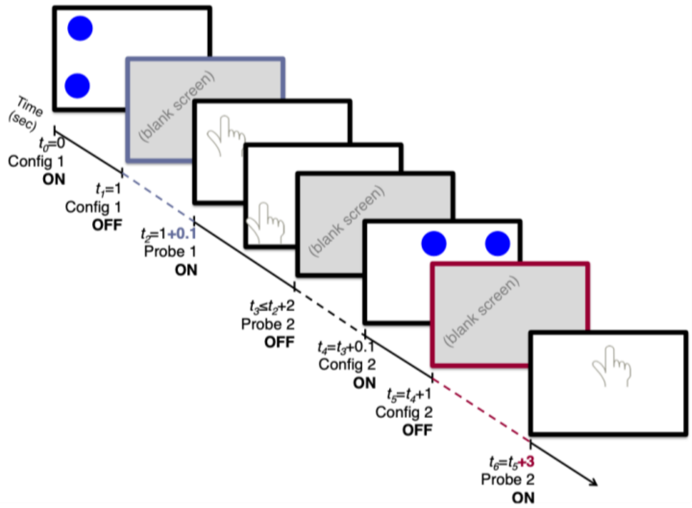
Table B3 reports the differences by child and household characteristics. As expected, accuracy was higher among older children. In the Peruvian sample, the average deviation declined by 5 per cent among index children compared to their younger siblings. Similarly, in the Ethiopian sample, the average deviation declined by 3 per cent among index children compared to their younger siblings. In both countries, the average deviation increased when the respondent was female, although the relationship was more established in Peru. Looking at household characteristics, in both Peru and Ethiopia, the average deviation was lower for children from the top wealth quintile (by 6 per cent and 3 per cent, respectively) and for children from urban areas (by 5 per cent and 2 per cent, respectively).

4.3. ‘Finding the dots’ game: measuring working memory

4.3.1. Game description

The working memory assessment is based on the ‘delayed match to sample ’paradigm established in cognitive neuroscience and cognitive psychology (Thomason et al. 2009). This game is a spatial working memory task consisting of 42 trials. Figure 8 provides a visualisation of how the task appears to the respondent.

Figure 8. *Long and short delay screens for the 'finding the dots' game*



In each trial, the child is shown a screen that has one dot (low-stimuli trials) or two to three dots on it (high stimuli trials). The child has 2 seconds to look at the screen and mentally take note of where the dots are located. After those 2 seconds are over, the screen goes blank, and the child must wait for a time while holding in mind the locations of the dots. Waiting time will be for either 0.1 seconds (short-delay trials) or 3 seconds (long-delay trials). After the waiting time, the screen 'opens', and the child must touch the screen as close as possible to where the dots used to be. The game includes 21 short-delay trails (seven low-stimuli and 14 high-stimuli trials) and 21 long-delay trails (seven low-stimuli and 14 high-stimuli trials) (see Table 6).

Table 6. *Number of trials in the 'finding the dots' game*

		Stimuli Low (1 dot)	Stimuli High (2-3 dots)	Total number of trials
Delay	Short (0.1 sec)	7	14	21
	Long (3 secs)	7	14	21
Total number of trials		14	28	42

4.3.2. Performance measure

Working memory is the ability to hold in mind and manipulate stimuli that are no longer present in the environment. This is a primary executive function (Miyake et al. 2000). While this is a simple cognitive function, it is a necessary component of many more complex abilities such as high-level reasoning, planning, or language comprehension. Children perform better on working memory tasks as they get older, and both child and adults recruit the prefrontal cortex when performing working memory tasks (Thomason et al. 2009). Working memory ability in childhood is linked with performance in school even after controlling for content of knowledge (Blair 2002), and training of working memory and executive function more generally is associated with decreased behavioural problems and increased academic performance (Diamond et al. 2007; Klingberg 2010).

In each trial of this game, the child's performance is measured as the Euclidean distance (in pixels) of their touch to the centre of the dot they were shown (in log-scale for empirical analysis). This represents a continuous measure of accuracy; it reflects not only whether a child could hold in mind the location of the dot, but also how precisely they could hold that location in mind.

Two sets of baseline/challenge measures can be computed on this task. On high-stimuli trials (screens with two or three dots – higher memory load), respondents must mentally rehearse to themselves a larger amount of spatial information relative to low-stimuli trials (screen with one dot – lower memory load). Similarly, on long-delay trials (3 second delay) they must continue the mental rehearsal for a longer time than on short-delay trials (0.1 second delay). High-stimuli and long-delay trials can therefore be considered as challenge trials and low-stimuli and short-delay trials as baseline trails. The distributions of the performance indicator described above are presented in Figures 9 and 10 for both short/long delay and low/high stimuli, respectively.

Overall, children in both countries are more accurate in their response in short-delay trials and in low-stimuli trails than in long-delay trials and in high-stimuli trials. On average, short-delay deviations were 4.23 for Ethiopia and 4.22 for the Peru sample, while long-delay deviations were 4.51 and 4.44, respectively. On the other hand, low-stimuli deviations were 3.89 and 3.87 while high-stimuli deviations were 4.46 and 4.41 for Ethiopia and Peru, respectively.

Figure 9. *Performance indicator distributions based on delay: Euclidean distances from touch to correct location (log scale), Ethiopia and Peru index children*

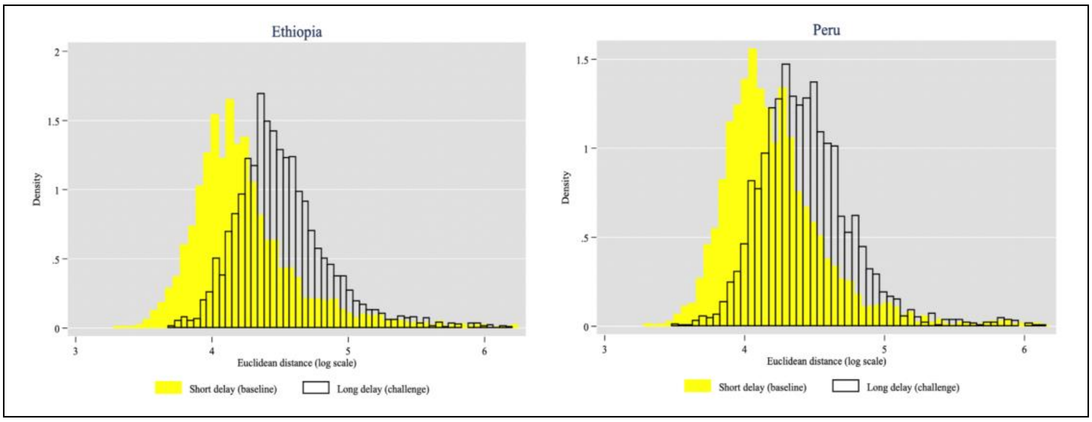
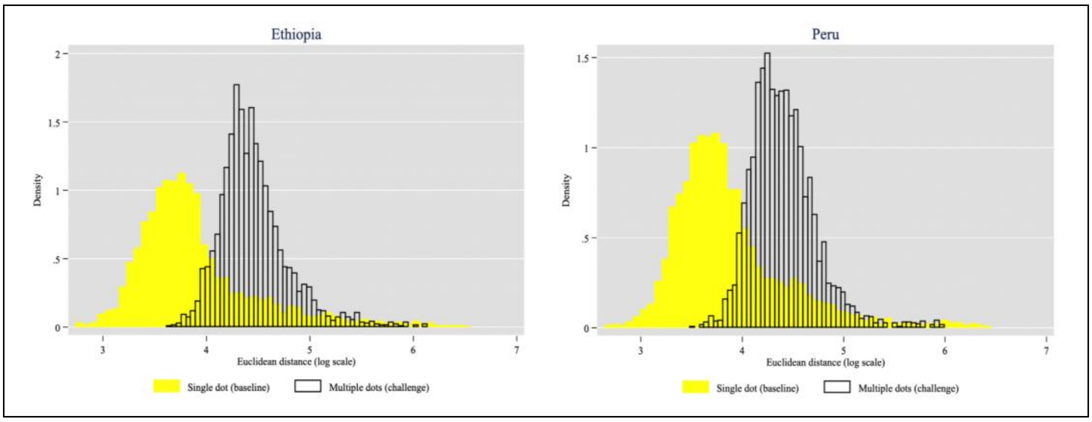


Figure 10. *Performance indicator distributions based on number of stimuli: Euclidean distances from touch to correct location (log scale), Ethiopia and Peru index children*



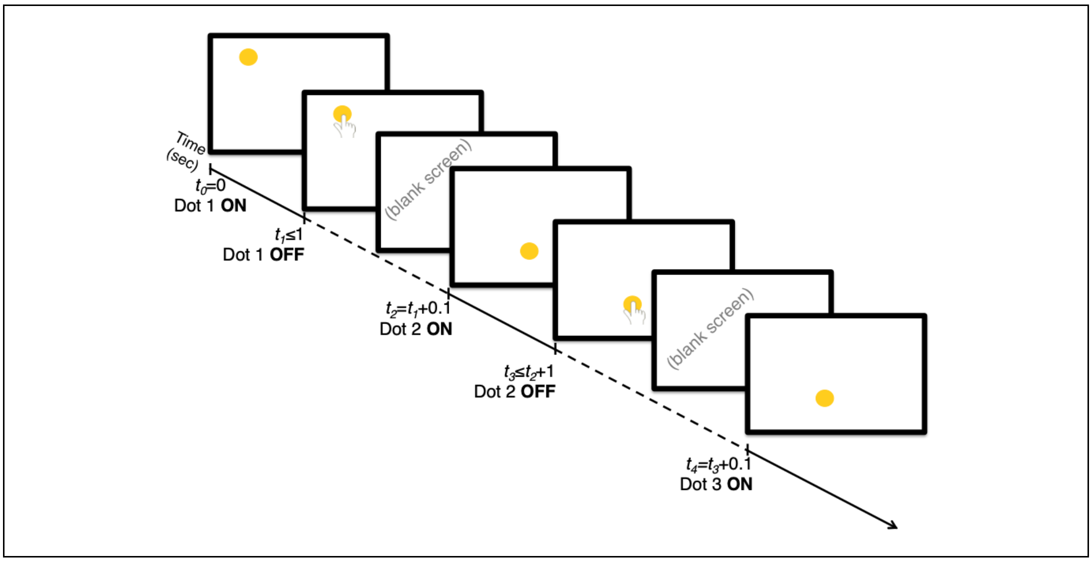
Tables B4 and B5 report how the average Euclidean distance differed by child and household characteristics, for long- versus short-delay trials and for multiple-dot versus single-dot trials, respectively. As expected, for both sets of measures, the average deviation was lower for older children. In both measures, the average deviation was also lower for males (which was not a priori expected). When comparing household characteristics, we found that accuracy was higher among children from wealthier households and children from urban areas. Accuracy was also consistently higher with more maternal schooling attainment, though this difference was less established in Ethiopia than in Peru. These findings again replicate previous observations using working memory tasks in high-income countries (Finn et al. 2010; Kharitonova, Winter, and Sheridan 2015; Noble, McCandliss, and Farah 2007).

4.4. 'Catching chickens' game: measuring implicit learning

4.4.1. Game description

The implicit learning game is presented to the respondents as a game called 'catching chickens' (or 'chasing the dots'). It is based on the 'serial reaction time 'paradigm, which is standard in cognitive science (Lum, Ullman, and Conti-Ramsden 2013; Nissen and Bullemer 1987; Ruitenberg, Verwey, and Abrahamse 2015). One hundred and seventy-five small yellow dots are presented one at a time in rapid succession. The child's task is to touch each dot as quickly as possible before it disappears. Each presentation of a dot is a trial, and each succession of 35 dots is referred to as a 'block'. The dot appears in one of four locations on the screen (the screen is divided into four quadrants, ABCD, from top left to bottom right), and disappears after 1 second or as soon as the child touched it (whichever comes first). Figure 11 provides a visualisation of how the task appears to the respondent.

Figure 11. *Screens in the 'catching chickens' game*



The succession of dot locations on the screen follows no pattern for the first 35 stimuli (block 1). However, in the next 70 stimuli (blocks 2 and 3), the dots are presented in 10 repeated cycles of seven locations ('ADABDDB'). The next 35 dots again follows no pattern (block 4), and the last 35 (block 5) are five more repeats of the ADABDDB pattern. The game moves seamlessly from block to block so the respondent would be unlikely to consciously note the patterned versus non-patterned movement. In total the child plays 70 non-patterned trials and 105 patterned trails (Table 7).

Table 7. *Number of trials in the 'catching chickens' game per block and per type*

Block	Type	Number of trials
1	Non-patterned	35
2	Patterned	35
3	Patterned	35
4	Non-patterned	35
5	Patterned	35
Total number of trials		175

4.4.2. Performance measure

Implicit learning is the ability of the body's motor system to recognise and respond to regularities in the environment, even when the person is not aware of these regularities. This ability is a very basic and primary form of learning, as it relies on the basal ganglia – deep brain structures which are conserved across species (Aron, Gluck, and Poldrack 2006). In this game, individuals press the dot more quickly when the movement of the dot follows a pattern, even when they themselves are unaware of the pattern (Pasupathy and Miller 2005). The ability to speed up with patterned presentations relative to random presentations, implicit learning, has been strongly linked with language acquisition in infancy and early childhood (Arnon 2019).

The key performance measure in this game is a respondent's reaction time – elapsed time from the appearance of the dot until the respondent touches it. There are two possible performance measures, depending on which baseline and challenge trials are selected in computing the performance indicator. As a first indicator, one can consider as baseline trials all the non-patterned blocks (blocks 1 and 4) while the challenge trials are all the patterned blocks (blocks 2, 3, and 5). As a second indicator, one can only consider a subset of non-patterned trials (block 4) and patterned trials (block 5) as baseline and challenge trials, respectively. While this subset decreases the number of trials per child in the sample, it lets the researcher focus on trials that occurred after the respondents have had time to learn the pattern. In both cases, the speed advantage that a respondent acquires from having implicitly learnt the pattern reflects their capacity to learn without conscious awareness. The speed advantage is assessed based on the difference in a respondent's mean response time on challenge trials versus baseline trials.

Figure 12 illustrates the reaction time distributions in the Ethiopian and Peruvian samples for both baseline (non-patterned) and challenge (patterned) trials when using the first definition. On average, the reaction time of Ethiopian and Peruvian children to non-patterned trials was 0.74 seconds. The reaction time to patterned trials of Peruvian children was lower than that of Ethiopian children (0.71 for Peruvian children compared to 0.72 for Ethiopian children). Figure 13 shows the reaction time distributions of both countries when adopting the second definition. On average, the reaction time of Ethiopian and Peruvian children to block 4 non-patterned trials was 0.73 seconds. Peruvian children's reaction to time block 5 patterned trials was 0.70 seconds, while Ethiopian children's reaction time to the same patterned trials was 0.71 seconds.

Figure 12. *Performance indicator distributions based on all non-patterned and patterned blocks: reaction time (in seconds), Ethiopia and Peru Younger Cohort index children*

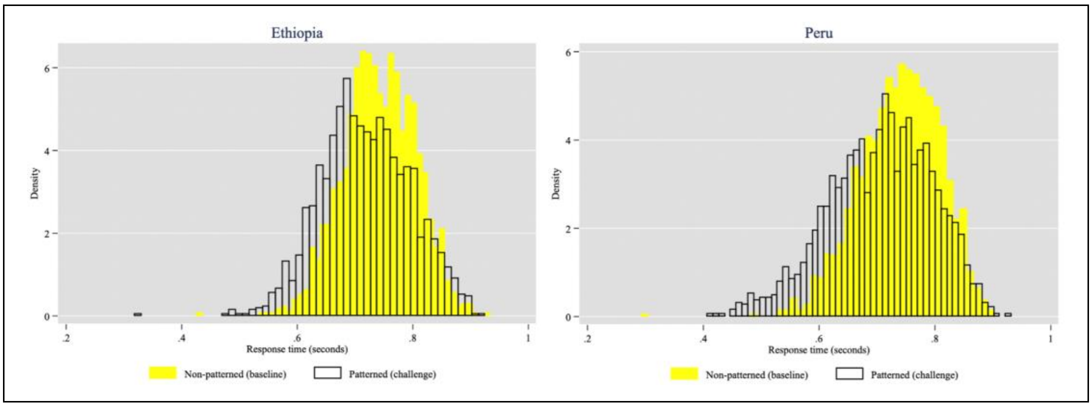


Figure 13. *Performance indicator distribution based on block 4 (non-patterned) and block 5 (patterned): reaction time (in seconds), Ethiopia and Peru Younger Cohort index children*

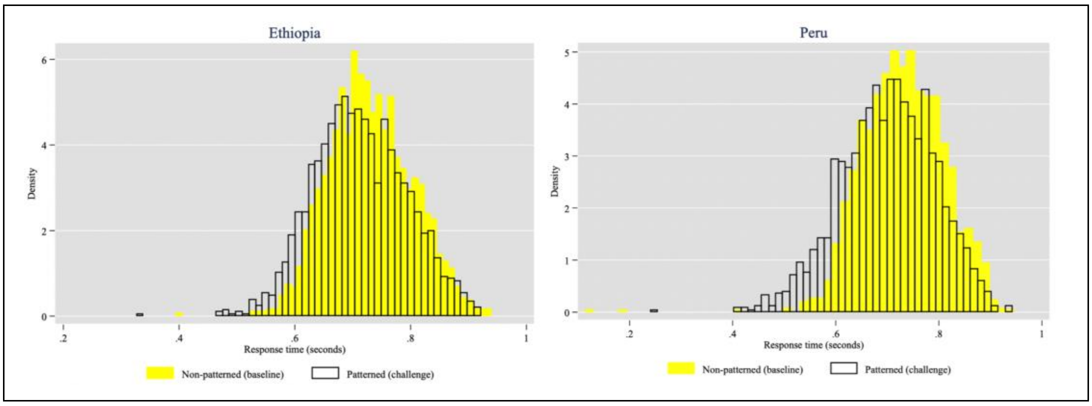


Table B6 reports the differences by child and household characteristics based on all non-patterned (baseline) and patterned (challenge) blocks. As expected, average response times were significantly lower among older children. On average, index children in Peru and Ethiopia had response times that were 0.06 and 0.05 seconds lower than their younger siblings, respectively. Males also had lower average response times than females in both countries. When analysing household characteristics, we observe that wealthier children and children from urban areas performed better on average. Children from the top wealth quintile in Peru and Ethiopia had response times that were 0.038 and 0.034 seconds quicker than those in the lowest wealth quintile, respectively, while children from urban areas reacted 0.029 and 0.025 seconds faster, respectively. Response times were also lower for children whose mothers have more schooling. Fewer studies have examined individual differences in implicit learning than in other aspects of cognitive function assessed in RACER; however, research does suggest that family experiences and age impact implicit learning task performance (Finn et al. 2019; Sheridan et al. 2018).

5. Tips for effective use of RACER data in statistical analysis

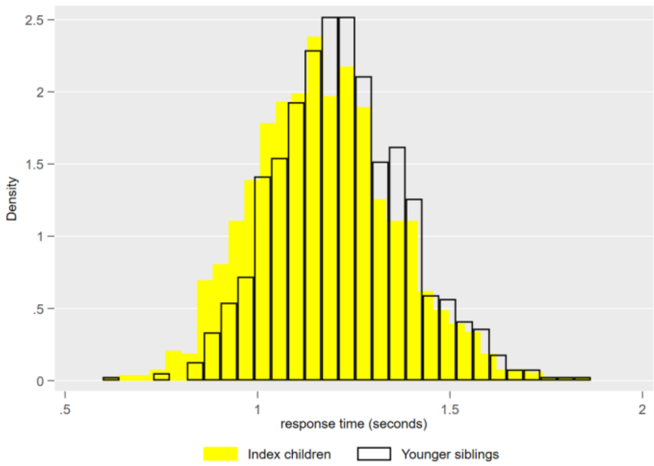
5.1 Controlling for RACER baseline trial measures

During analysis, it is recommended that challenge performance measures are always compared to baseline performance measures, so that the within-person difference is used as an indication of the person's foundational cognitive ability specifically, rather than their general ability to perform cognitive tasks. This comparison of these two types of trials should help account for any between-country, cross-children, and setting differences in task performance related to interpreting two-dimensional stimuli or responding by touching a tablet, as these should contribute equally to both baseline and challenge trials.

5.2 Controlling for age

In order to compare assessed cognitive skills between index children and their siblings, it is recommended to control for the effect of age on the development of foundational cognitive skills. It would be inappropriate to simply estimate differences in foundational cognitive skills across these groups by comparing the raw task data, because of imperfect overlap in the performance distributions due to differences in age. For example, as Figure 14 illustrates for the Ethiopian sample, the distributions of response times in baseline trials on the inhibitory control game differed substantially for index children and their younger siblings. The median response time on baseline trails among younger siblings was 1.21 seconds; among the index children, this response time was at the 70th percentile.

Figure 14. *Baseline indicator distribution for index children versus younger siblings: response time (in seconds), Ethiopia*



If challenge effects vary across the baseline distributions, the difference between average challenge effects among the index children and the average effects among their siblings will in part reflect differences in the distribution of baseline response times between ages, rather than differences in foundational cognitive skills. Therefore, to be able to pool observations and compare cognitive outcomes across different ages, we recommend controlling for the effect of age on foundational cognitive skills (or, alternatively, to standardise by age).

As an example, the Stata code below shows how to standardise the performance measure for long-term memory (*task1*) according to age in years (for individuals aged 6 to 26) and generate a new, standardised challenge measure:

```
forval i = 6/26{
        egen std_task1_`i’ =std(task1) if age==`i’
}


gen std_task1=.

forval i = 6/26{
replace std_task1 =std_task_`i’ if age==`i’
}


drop std_task1_*
```

5.3 Other aspects to consider

A team of researchers that has analysed the RACER data collected in the Young Lives study, as part of the ‘Early-life determinants of foundational cognitive skills: the roles of nutritional investments, pre-schooling and anti-poverty social programs’[11] project funded by the National Institutes of Health (the YL–NIH team), decided to take the following steps for statistical analysis of the RACER data:

i **Inverting measures for improved communication**: As the relevant performance indicators differ between the RACER tasks, the communication of improvements in foundational cognitive skill is complicated. For example, in the inhibitory control task, one of the possible performance measures is response/reaction time (in seconds) – in which a lower score reflects higher inhibitory control. Conversely, in the long-term memory game, the preferred performance measure is the percentage of trials the child got correct at first touch – in which a higher score reflects higher long-term memory. In order to make the communication of statistical analysis more straightforward, the YL–NIH team inverted the baseline and performance measures of the inhibitory control, working memory and implicit learning games so that, for all RACER tasks, an increase in the outcome variable is

[11] The publications from this project can be found at https://www.younglives.org.uk/research-project/cognitive-skills-ethiopia-and-peru

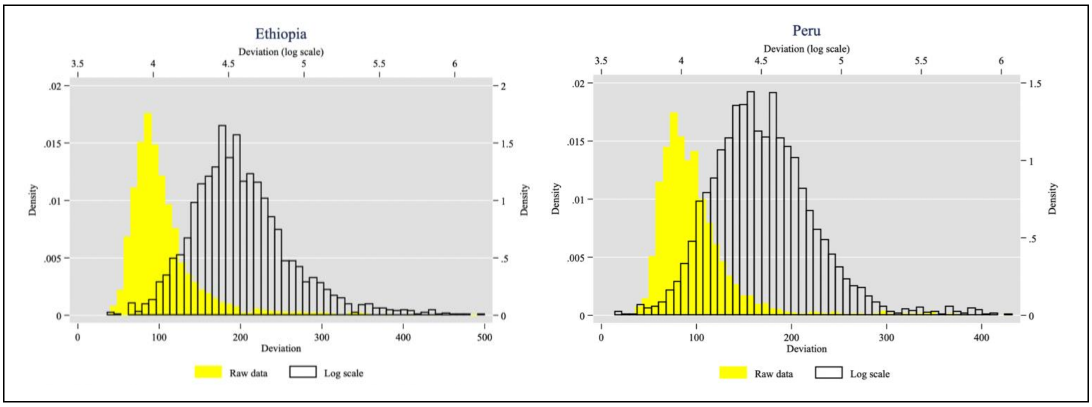
associated with an improvement in foundational cognitive skills. For example, the Stata code below shows how the researchers inverted the challenge measure of working memory (*task3*):

```
gen itask3 = 1/task3
```

ii **Collapsing trial-level data into child-level data**: The YL–NIH team decided to collapse the trial-level data into child-level data. This implies that, for each child, for each task there is one baseline observation and one challenge observation, each of which is the average of multiple trials.

iii **Taking the logarithm of Euclidean distance measures**: The YL–NIH team found that, for performance outcomes which measure the Euclidean distance (in pixels), taking the logarithm leads to a distribution that is much closer to approximating the normal distribution than do the raw data. As an illustration, Figure 15 compares, for game three ('finding the dots'), the distribution of the raw Euclidean distance (in pixels) data and its logarithm. The logarithms are included in the datasets that are being archived (Section 6).

Figure 15. *Performance indicator distributions, raw values versus logarithm: Euclidean distances (in pixels) based on multiple-dot, long-delay trials, Ethiopia and Peru*



i **Definition of challenge and baseline trials for working memory**: For working memory there are two types of baseline/challenge comparisons: single dot versus multiple dots, and short-delay versus long-delay trials (see Section 4). For their analysis, the YL–NIH team decided to define baseline trials as those trials that were simultaneously short delay and single dot, and challenge trials as those that were simultaneously long delay and multiple dots.

ii **Definition of long-term memory and inhibitory control performance measures**: In the games for long-term memory and inhibitory control, there are two potential performance measures available – both of which capture valuable information about the child's foundational cognitive skills (see Table 8). For long-term memory, the YL–NIH team chose to use the percentage of trials correct at first touch as the preferred measure. For inhibitory control, the team chose to create an equally weighted average of both performance standardised indicators as the preferred outcome of interest. The Stata code for the challenge performance measure is shown below:

```
g task2_a=0.5*euclideandave_opp +0.5*resptimeave_opp
```

Table 8. *Potential performance measures in RACER games*

RACER game	Cognitive task	Possible outcome measures	Preferred outcome measure
Memory game	Long-term memory	Percentage of trials correct at first touch Total number of touches/choices made before correct pair	Percentage of trials correct at first touch
Sides game	Inhibitory control	Response time (in seconds) Euclidean distance of touch from centre of dot (in pixels) (in log scale)	Equally weighted average of response time and Euclidean distance
Finding the dots	Working memory	Euclidean distance of touch to centre of dot (in pixels) (in log scale)	Euclidean distance
Catching chickens/Chasing dots	Implicit learning	Response time (in seconds)	Response time

vi The YL–NIH team **standardised all outcomes** (after inversion and/or after taking logs).

6. Database and data dictionary

One dataset will be archived, which contains information on all RACER games in both Ethiopia and Peru. In the dataset, there is an identifier variable for the RACER task (task_num), which can be used to isolate a particular game. Tables C1 to C4 report a list of the variables available in the dataset for each foundational skill task. The averaged variables, constant at the child level, are obtained by averaging the data for each individual at the trial level.

The archived RACER dataset will be available for download from the UK Data Archive by the end of 2023 (Table 9).

Table 9. *List of archived FCS tasks and data dictionaries*

Task number	Data dictionary	Foundational cognitive skill tested
1 and 5	Table C1	Long-term memory
2	Table C2	Inhibitory control
3	Table C3	Working memory
4	Table C4	Implicit learning

References

Arnon, I. (2019) 'Statistical Learning, Implicit Learning, and First Language Acquisition: A Critical Evaluation of Two Developmental Predictions', *Topics in Cognitive Science* 11(3): 504–19.

Aron, A. M.A. Gluck, and R.A. Poldrack (2006) 'Long-term Test-retest Reliability of Functional MRI in a Classification Learning Task', *Neuroimage* 29(3): 1000–6.

Barkley, R.A., G. Grodzinsky, and G.J. DePaul (1992) 'Frontal Lobe Functions in Attention Deficit Disorder With and Without Hyperactivity: A Review and Research Report', *Journal of Abnormal Child Psychology* 20:163–188.

Bechara, A., D. Tranel, H. Damasio, R. Adolphs, C. Rockland, and A.R. Damasio (1995) 'Double Dissociation of Conditioning and Declarative Knowledge Relative to the Amygdala and Hippocampus in Humans', *Science* 269(5227): 1115–8.

Blair, C. (2002) 'School Readiness: Integrating Cognition and Emotion in a Neurobiological Conceptualization of Children's Functioning at School Entry', *American Psychologist* 57(2): 111–27.

Blair, C., and R.P. Razza (2007) 'Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten', *Child Development* 78(2): 647–63.

Briones, K. (2017) *'How Many Rooms Are There in Your House?' Constructing the Young Lives Wealth Index*, Young Lives Technical Note 43, Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-TN43_0.pdf (accessed 7 September 2022).

Chen, A., C. Panter-Brick, K. Hadfield, R. Dajani, A. Hamoudi, and M.A. Sheridan (2019) 'Minds Under Siege: Cognitive Signatures of Poverty and Trauma in Refugee and Non-Refugee Adolescents', *Child Development* 90(6): 1856–65.

Cueto, S., J. León, G. Guerrero, and I. Muñoz (2009) Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 2 of Young Lives, Young Lives Technical Note 15, Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-TN15-Cueto-Psychometric-Characteristics.pdf (accessed 7 September 2022).

Cueto S., and J. León (2012) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives*, Young Lives Technical Note 25, Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-TN25_Cueto.pdf (accessed 7 September 2022).

Diamond, A., W.S. Barnett, J. Thomas, and S. Munro (2007) 'Preschool Program Improves Cognitive Control', *Science* 318(5855): 1387–8.

Eichenbaum, H. (2004) 'Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory', *Neuron* 44: 109–20.

Escobal, J., and E. Flores (2008) *An Assessment of the Young Lives Sampling Approach in Peru*, Young Lives Technical Note 3, Oxford: Young Lives.

https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-TN3-Escobal-Sampling-Approach-In-Peru.pdf (accessed 7 September 2022).

Farah, M.J., D.M. Shera, J.H. Savage, L. Betancourt, J.M. Giannetta, N.L. Brodsky, E.K. Malmud, and H. Hurt (2006) 'Childhood Poverty: Specific Associations with Neurocognitive Development', *Brain Research* 1110: 166–74.

Favara, M., G. Crivello, M. Penny, C. Porter, E. Revathi, A. Sánchez, D. Scott, L.T. Duc, T. Woldehanna, and A. McKay (2021) 'Cohort Profile Update: The Young Lives Study', *International Journal of Epidemiology* 50(6): 1784–5e.

Finn, A.S., M.A. Sheridan, C.L.H. Kam, S. Hinshaw, and M. D'Esposito (2010) 'Longitudinal Evidence for Functional Specialization of the Neural Circuit Supporting Working Memory in the Human Brain', *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30(33): 11062–7.

Finn, A.S., M. Kharitonova, N. Holtby, and M.A. Sheridan (2019) 'Prefrontal and Hippocampal Structure Predict Statistical Learning Ability in Early Childhood', *Journal of Cognitive Neuroscience* 31(1): 126–37. doi:10.1162/jocn_a_01342.

Ford, C.B., H.Y. Kim, L. Brown, J.L. Aber, and M.A. Sheridan (2019) 'A Cognitive Assessment Tool Designed for Data Collection in the Field in Low- and Middle-income Countries', *Research in Comparative and International Education* 14(1): 141–57.

Gabrieli, J.D. (1993) 'Disorders of Memory in Humans', *Current Opinions in Neurology and Neurosurgery* 6(1): 93–7.

Ghetti, S., and L. Angelini (2008) 'The Development of Recollection and Familiarity in Childhood and Adolescence: Evidence from the Dual-process Signal Detection Model', *Child Development* 79(2): 339–58.

Grammer, J., J.L. Coffman, and P. Ornstein (2013) 'The Effect of Teachers' Memory-Relevant Language on Children's Strategy Use and Knowledge', *Child Development* 84(6): 1989–2002.

Hackman, D.A., and M.J. Farah (2009) 'Socioeconomic Status and the Developing Brain', *Trends in Cognitive Sciences* 13(2): 65–73.

Hamoudi, A., and M.A. Sheridan (2015) 'Unpacking the Black Box of Cognitive Ability: A Novel Tool for Assessment in a Population-based Survey', Mimeo.

Hannula, D.E., D. Tranel, and N.J. Cohen (2006) 'The Long and the Short of it: Relational Memory Impairments in Amnesia, Even at Short Lags', *Journal of Neuroscience* 26(32): 8352–9.

Hanson, J., A. Chandra, B. Wolfe, and S. Pollak (2011) 'Association Between Income and the Hippocampus', *PLoS ONE* 6(5): e18712. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018712 (accessed 7 September 2022).

Kharitonova, M., W. Winter, and M.A. Sheridan (2015) 'As Working Memory Grows: A Developmental Account of Neural Bases of Working Memory Capacity in 5- to 8-Year-Old Children and Adults', *Journal of Cognitive Neuroscience* 27(9): 1175–88.

Kiely, K. (2014) 'Cognitive Function' in K.M. Michalos (ed.) *Encyclopedia of Quality of Life and Well-Being Research*, 974–8, Dordrecht: Springer.

Kim, J.J., and K.S. Yoon (1998) ‘Stress: Metaplastic Effects in the Hippocampus’, *Trends in Neurosciences* 21(120): 505–9.

Kim, H.Y., L. Brown, C.T. Dolan, M.A. Sheridan, and J.L. Aber (2020) ‘Post-migration Risks, Developmental Processes, and Learning among Syrian Refugee Children in Lebanon’, *Journal of Applied Developmental Psychology*, 69: 101142.

Klingberg, T. (2010) ‘Training and Plasticity of Working Memory’, *Trends in Cognitive Sciences* 14: 317–24.

Lum, J.A., M.T. Ullman, and G. Conti-Ramsden (2013) ‘Procedural Learning is Impaired in Dyslexia: Evidence from a Meta-analysis of Serial Reaction Time Studies’, *Research in Developmental Disabilities* 34(10): 3460–76.

McEwen, B.S. (2001) ‘Plasticity of the Hippocampus: Adaptation to Chronic Stress and Allostatic Load’, *Annals of the New York Academy of Sciences* 933: 265–77.

Miyake, A., N.P. Friedman, M.M. Emerson, A.H. Witzki, A. Howerter, and A.D. Wager (2000) ‘The Unity and Diversity of Executive Functions and their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis’, *Cognitive Psychology* 41(1): 49–100.

Miyake, A., and N.P. Friedman (2012) ‘The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions’, *Current Directions in Psychological Science* 21(1): 8–14.

Nissen, M.J., and P. Bullemer (1987) ‘Attentional Requirements of Learning: Evidence from Performance Measures’, *Cognitive Psychology* 19(1):1–32.

Noble, K.G., B.D. McCandliss, and M.J. Farah (2007) ‘Socioeconomic Gradients Predict Individual Differences in Neurocognitive Abilities’, *Developmental Science* 10: 464–80.

Outes-Leon, I., and A. Sánchez (2008) *An Assessment of the Young Lives Sampling Approach in Ethiopia*, Young Lives Technical Note 1, Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-TN1-OutesLeon-Sampling-Approach-In-Ethiopia.pdf (accessed 7 September 2022).

Pasupathy, A., and E.K. Miller (2005) ‘Different Time Courses of Learning-related Activity in the Prefrontal Cortex and Striatum’, *Nature* 433(7028): 873–6.

Ruitenberg, M.F.L., W.B. Verwey, and E.L. Abrahamse (2015) ‘What Determines the Impact of Context on Sequential Action?’, *Human Movement Science* 40: 298–314.

Sheridan, M., M. Kharitonova, R.E. Martin, A. Chatterjee, and J.D.E. Gabrieli (2014) ‘Neural Substrates of the Development of Cognitive Control in Children Ages 5–10 Years’, *Journal of Cognitive Neuroscience* 26(8): 1840–50.

Sheridan, M.A., K.A. McLaughlin, W. Winter, N. Fox, C. Zeanah, and C.A. Nelson (2018) ‘Early Deprivation Disruption of Associative Learning is a Developmental Pathway to Depression and Social Problems’, *Nature Communications* 9: Article 2216.

Shonkoff, J.P., W.T. Boyce, and B.S. McEwen (2009) ‘Neuroscience, Molecular Biology, and the Childhood Roots of Health Disparities: Building a New Framework for Health Promotion and Disease Prevention’, *JAMA* 301(21): 2252–9.

Simon, J.R., and A.P. Rudell (1967) ‘Auditory S-R Compatibility: The Effect of an Irrelevant Cue on Information Processing’, *Journal of Applied Psychology* 51(3): 300–4.

Skinner, E.A., M.J. Zimmer-Gembeck, J.P. Connell, J.S. Eccles, and J.G. Wellborn (1998) 'Individual Differences and the Development of Perceived Control', *Monographs of the Society for Research in Child Development* 63(2-3): Serial No. 254.

Thomason, M.E., E. Race, B. Burrows, S. Whitfield-Gabrieli, G.H. Glover, and J.D. Gabrieli (2009) 'Development of Spatial and Verbal Working Memory Capacity in the Human Brain', *Journal of Cognitive Neuroscience* 21(2):316–32.

Young Lives (2017) 'A Guide to Young Lives Research', Oxford: Young Lives. http://www.younglives.org.uk/sites/www.younglives.org.uk/files/GuidetoYLResearch_0.pdf (accessed 7 September 2022).

Appendices

Appendix A

Figure A1. *Shapes used in the memory game*

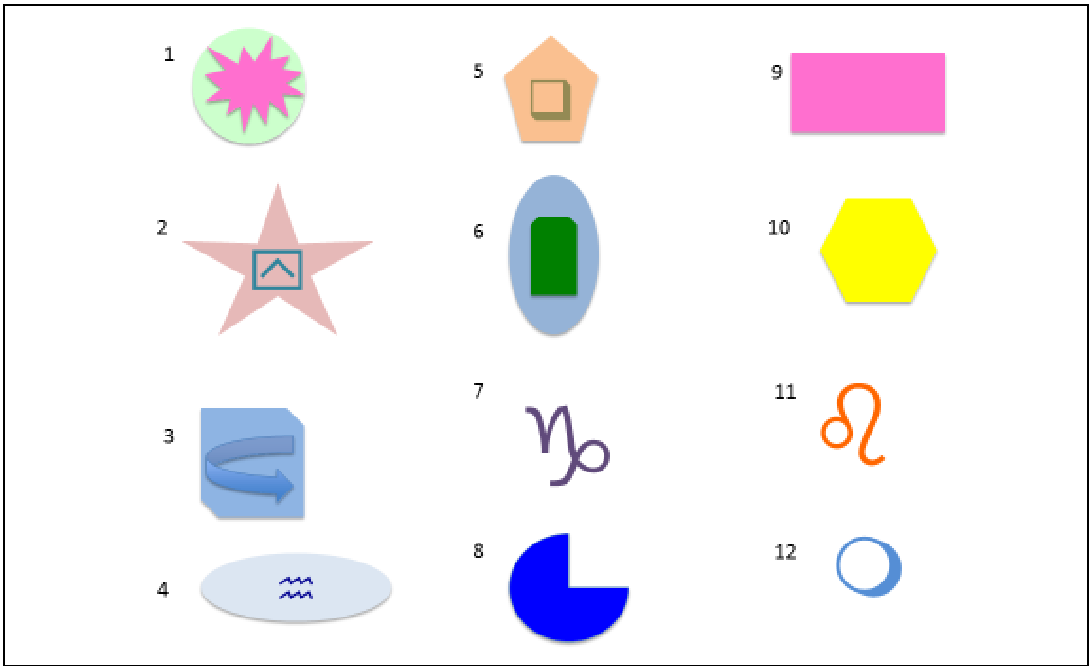
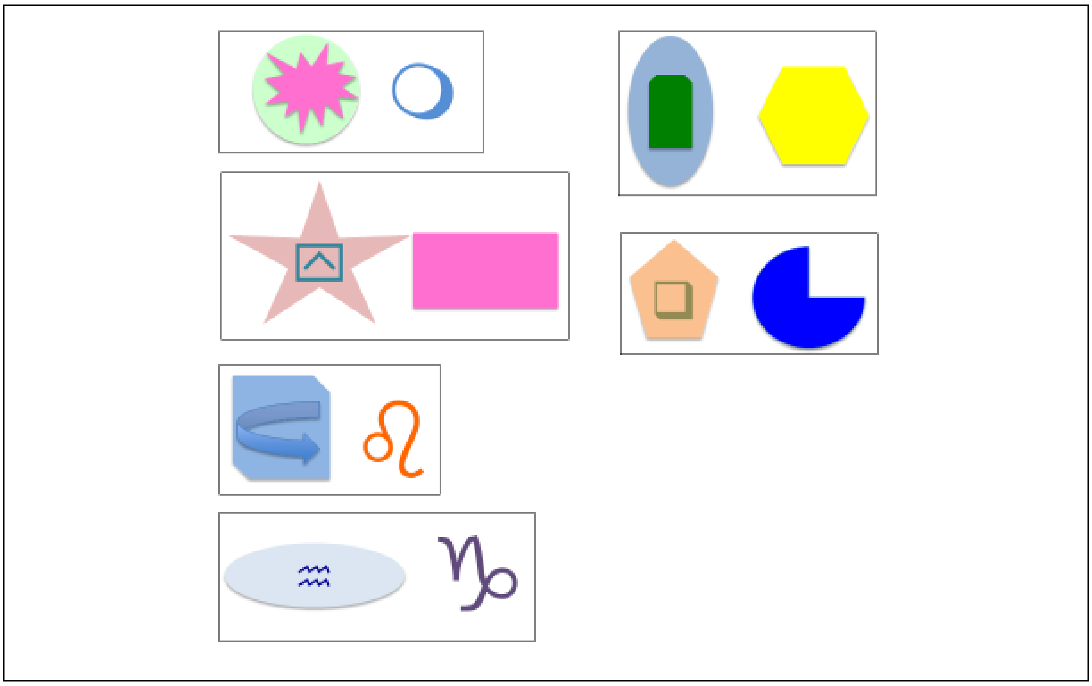


Figure A2. *Pairs used in the memory game*



Appendix B

Table B1. *Long-term memory (average percentage of correct answers at the first touch)*

Peru	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,873	0.408		0.220	
Younger sibling	783	0.366	0.000	0.244	0.002
By age					
7 or less	217	0.351		0.251	
8	171	0.332	0.201	0.234	0.369
9	239	0.379	0.001	0.248	0.833
10	130	0.414	0.000	0.242	0.652
11	1,001	0.409	0.000	0.216	0.006
12	860	0.405	0.000	0.225	0.059
Index child only					
By gender					
Male	944	0.406		0.218	
Female	929	0.410	0.562	0.222	0.632
By wealth (Round 1)					
Bottom quintile	375	0.371		0.222	
Top quintile	371	0.445	0.000	0.218	0.759
By urban–rural					
Urban	1,375	0.421		0.219	
Rural	497	0.369	0.000	0.224	0.622
By maternal schooling attainment					
Less than primary	650	0.377		0.218	
Complete primary	526	0.403	0.004	0.214	0.620
Complete secondary	377	0.430	0.000	0.227	0.458
Higher education	255	0.463	0.000	0.229	0.420
By maternal native tongue					
Non-Spanish	529	0.387		0.211	
Spanish	1,314	0.416	0.001	0.225	0.117

Ethiopia	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,803	0.365		0.270	
Younger sibling	933	0.341	0.000	0.298	0.000
By age					
7 or less	125	0.345		0.323	
8	242	0.328	0.320	0.306	0.485
9	314	0.335	0.516	0.281	0.057
10	193	0.342	0.842	0.288	0.155
11	639	0.359	0.370	0.281	0.050
12	1,206	0.371	0.097	0.267	0.005
13	2	0.388	0.724	0.250	0.656
Index child only					
By gender					
Male	951	0.374		0.268	
Female	852	0.355	0.013	0.271	0.787
By wealth (Round 1)					
Bottom quintile	373	0.351		0.311	
Top quintile	351	0.431	0.000	0.270	0.011
By urban–rural					
Urban	664	0.416		0.260	
Rural	1,138	0.336	0.000	0.276	0.123
By maternal schooling attainment					
No schooling	877	0.340		0.257	
Lower primary (grades 1-4)	417	0.362	0.013	0.292	0.004
Upper primary (grades 5-8)	274	0.406	0.000	0.280	0.097
More than 8 grades	174	0.445	0.000	0.280	0.171
By region					
Addis Ababa	263	0.460		0.278	
Amhara	351	0.362	0.000	0.230	0.003
Oromia	384	0.378	0.000	0.302	0.149
SNNP	421	0.351	0.000	0.317	0.024
Tigray	380	0.306	0.000	0.217	0.000

Notes: For each child, trial-level data were collapsed into a unique observation (the average). P-values correspond to the results of t-tests testing the null hypothesis that the mean difference is equal to zero.

Table B2. *Inhibitory control (average response time, in seconds)*

Peru	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,877	1.114		1.076	
Younger sibling	785	1.284	0.000	1.224	0.000
By age					
7 or less	271	1.363		1.287	
8	172	1.273	0.000	1.227	0.002
9	240	1.260	0.000	1.202	0.000
10	130	1.218	0.000	1.161	0.000
11	1,003	1.120	0.00	1.083	0.000
12	862	1.106	0.000	1.069	0.000
Index child only					
By gender					
Male	946	1.094		1.056	
Female	931	1.133	0.000	1.097	0.000
By wealth (Round 1)					
Bottom quintile	375	1.178		1.137	
Top quintile	373	1.033	0.000	1.000	0.000
By urban–rural					
Urban	1,378	1.091		1.053	
Rural	498	1.177	0.000	1.141	0.000
By maternal schooling attainment					
Less than primary	650	1.162		1.124	
Complete primary	528	1.113	0.000	1.078	0.000
Complete secondary	377	1.081	0.000	1.042	0.000
Higher education	256	1.035	0.000	0.999	0.000
By maternal native tongue					
Non-Spanish	528	1.144		1.109	
Spanish	1,318	1.102	0.000	1.064	0.000

Ethiopia	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,811	1.175		1.133	
Younger sibling	938	1.272	0.000	1.216	0.000
By age					
7 or less	125	1.312		1.266	
8	243	1.293	0.316	1.233	0.096
9	316	1.268	0.016	1.208	0.002
10	195	1.259	0.008	1.199	0.001
11	641	1.185	0.000	1.140	0.000
12	1,212	1.170	0.000	1.129	0.000
13	2	1.136	0.124	1.060	0.117
Index child only					
By gender					
Male	957	1.173		1.130	
Female	854	1.178	0.574	1.136	0.378
By wealth (Round 1)					
Bottom quintile	375	1.185		1.142	
Top quintile	354	1.113	0.000	1.088	0.000
By urban–rural					
Urban	669	1.145		1.110	
Rural	1,141	1.193	0.000	1.146	0.000
By maternal schooling attainment					
No schooling	879	1.203		1.158	
Lower primary (grades 1-4)	418	1.167	0.000	1.122	0.000
Upper primary (grades 5-8)	279	1.145	0.000	1.107	0.000
More than 8 grades	174	1.093	0.000	1.067	0.000
By region					
Addis Ababa	264	1.116		1.090	
Amhara	351	1.173	0.000	1.124	0.005
Oromia	387	1.177	0.000	1.125	0.004
SNNP	425	1.157	0.002	1.126	0.004
Tigray	380	1.238	0.000	1.187	0.000

Notes: For each child, trial-level data were collapsed into a unique observation (the average). P-values correspond to the results of t-tests testing the null hypothesis that the mean difference is equal to zero.

Table B3. *Inhibitory control (average Euclidean distance, log scale)*

Peru	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,877	4.439		3.207	
Younger sibling	785	4.648	0.000	3.414	0.000
By age					
7 or less	271	4.678		3.574	
8	172	4.640	0.446	3.410	0.001
9	240	4.617	0.199	3.337	0.000
10	130	4.639	0.488	3.271	0.000
11	1,003	4.522	0.000	3.207	0.000
12	862	4.564	0.009	3.204	0.000
Index child only					
By gender					
Male	946	4.489		3.214	
Female	931	4.589	0.000	3.199	0.391
By wealth (Round 1)					
Bottom quintile	375	4.668		3.272	
Top quintile	373	4.365	0.000	3.172	0.001
By urban–rural					
Urban	1,378	4.471		3.169	
Rural	498	4.727	0.000	3.310	0.000
By maternal schooling attainment					
Less than primary	650	4.663		3.287	
Complete primary	528	4.531	0.000	3.169	0.000
Complete secondary	377	4.421	0.000	3.165	0.000
Higher education	256	4.411	0.000	3.136	0.000
By maternal native tongue					
Non-Spanish	528	4.665		3.285	
Spanish	1,318	4.488	0.000	3.175	0.000

Ethiopia	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,811	4.598		3.381	
Younger sibling	938	4.745	0.000	3.564	0.000
By age					
7 or less	125	4.787		3.661	
8	243	4.780	0.895	3.599	0.186
9	316	4.771	0.758	3.568	0.039
10	195	4.675	0.043	3.499	0.000
11	641	4.608	0.000	3.400	0.000
12	1,212	4.592	0.000	3.373	0.000
13	2	4.771	0.964	3.421	0.428
Index child only					
By gender					
Male	957	4.578		3.388	
Female	854	4.619	0.084	3.373	0.365
By wealth (Round 1)					
Bottom quintile	375	4.624		3.418	
Top quintile	354	4.486	0.000	3.327	0.000
By urban–rural					
Urban	669	4.525		3.319	
Rural	1,141	4.640	0.000	3.418	0.000
By maternal schooling attainment					
No schooling	879	4.629		3.413	
Lower primary (grades 1-4)	418	4.602	0.373	3.365	0.017
Upper primary (grades 5-8)	279	4.506	0.000	3.332	0.001
More than 8 grades	174	4.592	0.377	3.337	0.007
By region					
Addis Ababa	264	4.497		3.299	
Amhara	351	4.554	0.157	3.351	0.062
Oromia	387	4.615	0.004	3.398	0.001
SNNP	425	4.565	0.085	3.426	0.000
Tigray	380	4.728	0.000	3.400	0.001

Notes: For each child, trial-level data were collapsed into a unique observation (the average). P-values correspond to the results of t-tests testing the null hypothesis that the mean difference is equal to zero.

Table B4. *Working memory (average Euclidean distance, log scale), long delay (challenge) versus short delay (baseline)*

Peru	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,872	4.383		4.150	
Younger sibling	780	4.586	0.000	4.370	0.000
By age					
7 or less	215	4.732		4.541	
8	170	4.535	0.000	4.342	0.000
9	239	4.546	0.000	4.304	0.000
10	130	4.488	0.000	4.253	0.000
11	999	4.394	0.000	4.154	0.000
12	861	4.369	0.000	4.146	0.000
Index child only					
By gender					
Male	941	4.337		4.105	
Female	931	4.429	0.000	4.196	0.000
By wealth (Round 1)					
Bottom quintile	374	4.485		4.269	
Top quintile	373	4.278	0.000	4.054	0.000
By urban–rural					
Urban	1,373	4.342		4.268	
Rural	498	4.497	0.000	4.108	0.000
By maternal schooling attainment					
Less than primary	650	4.466		4.234	
Complete primary	527	4.384	0.000	4.153	0.000
Complete secondary	376	4.315	0.000	4.070	0.000
Higher education	254	4.267	0.000	4.050	0.000
By maternal native tongue					
Non-Spanish	528	4.457		4.228	
Spanish	1,313	4.354	0.000	4.120	0.000

Ethiopia	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average	1,806	4.464		4.186	
Index child	935	4.691	0.000	4.424	0.000
Younger sibling					
By age	124	4.792		4.517	
7 or less	243	4.752	0.351	4.503	0.795
8	315	4.691	0.013	4.403	0.020
9	194	4.590	0.000	4.328	0.001
10	639	4.489	0.000	4.216	0.000
11	1,209	4.454	0.000	4.174	0.000
12	2	4.386	0.145	4.091	0.210
13					
Index child only					
By gender	953	4.425		4.150	
Male	853	4.508	0.000	4.227	0.000
Female					
By wealth (Round 1)	374	4.527		4.256	
Bottom quintile	352	4.393	0.000	4.125	0.000
Top quintile					
By urban–rural	666	4.407		4.131	
Urban	1,139	4.498	0.000	4.219	0.000
Rural					
By maternal schooling attainment	877	4.490		2.207	
No schooling	418	4.468	0.239	4.200	0.758
Lower primary (grades 1-4)	276	4.405	0.000	4.119	0.000
Upper primary (grades 5-8)	174	4.437	0.039	4.177	0.307
More than 8 grades					
By region	264	4.380		4.115	
Addis Ababa	351	4.431	0.021	4.148	0.186
Amhara	385	4.451	0.005	4.169	0.051
Oromia	422	4.534	0.000	4.253	0.000
SNNP	380	4.489	0.000	4.217	0.000
Tigray	1,806	4.464		4.186	

Notes: For each child, trial-level data were collapsed into a unique observation (the average). P-values correspond to the results of t-tests testing the null hypothesis that the mean difference is equal to zero.

Table B5. *Working memory (average Euclidean distance, log scale), multiple dots (challenge) versus single dot (baseline)*

Peru	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,872	4.347		3.794	
Younger sibling	780	4.550	0.000	4.057	0.000
By age					
7 or less	215	4.701		4.240	
8	170	4.511	0.000	4.019	0.001
9	239	4.497	0.000	4.008	0.000
10	130	4.450	0.000	3.914	0.000
11	999	4.356	0.000	3.797	0.000
12	861	4.336	0.000	3.788	0.000
Index child only					
By gender					
Male	941	4.297		3.781	
Female	931	4.397	0.000	3.807	0.246
By wealth (Round 1)					
Bottom quintile	374	4.446		3.964	
Top quintile	373	4.251	0.000	3.675	0.000
By urban–rural					
Urban	1,373	4.307		3.739	
Rural	498	4.456	0.000	3.945	0.000
By maternal schooling attainment					
Less than primary	650	4.425		3.912	
Complete primary	527	4.348	0.000	3.794	0.000
Complete secondary	376	4.279	0.000	3.680	0.000
Higher education	254	4.245	0.000	3.650	0.000
By maternal native tongue					
Non-Spanish	528	4.417		3.892	
Spanish	1,313	3.318	0.000	3.754	0.000

Ethiopia	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,806	4.411		3.822	
Younger sibling	935	4.633	0.000	4.155	0.000
By age					
7 or less	124	4.726		4.300	
8	243	4.694	0.450	4.291	0.907
9	315	4.627	0.013	4.117	0.010
10	194	4.538	0.000	4.018	0.000
11	639	4.438	0.000	3.848	0.000
12	1,209	4.400	0.000	3.810	0.000
13	2	4.351	0.168	3.524	0.113
Index child only					
By gender					
Male	953	4.371		3.797	
Female	853	4.455	0.000	3.850	0.031
By wealth (Round 1)					
Bottom quintile	374	4.468		3.943	
Top quintile	352	4.354	0.000	3.680	0.000
By urban–rural					
Urban	666	4.364		3.397	
Rural	1,139	4.439	0.000	3.895	0.000
By maternal schooling attainment					
No schooling	877	4.430		3.883	
Lower primary (grades 1-4)	418	4.417	0.488	3.837	0.154
Upper primary (grades 5-8)	276	4.357	0.000	3.695	0.000
More than 8 grades	174	4.403	0.278	3.721	0.000
By region					
Addis Ababa	264	4.340		3.677	
Amhara	351	4.379	0.073	3.764	0.016
Oromia	385	4.393	0.031	3.827	0.000
SNNP	422	4.477	0.000	3.911	0.000
Tigray	380	4.434	0.000	3.872	0.000

Notes: For each child, trial-level data were collapsed into a unique observation (the average). P-values correspond to the results of t-tests testing the null hypothesis that the mean difference is equal to zero.

Table B6. *Implicit learning (reaction times, in seconds)*

Peru	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,869	0.706		0.741	
Younger sibling	780	0.761	0.000	0.772	0.000
By age					
7 or less	214	0.776		0.773	
8	172	0.761	0.047	0.772	0.905
9	238	0.755	0.006	0.772	0.773
10	130	0.748	0.001	0.771	0.791
11	1,001	0.709	0.000	0.743	0.000
12	856	0.703	0.000	0.738	0.000
Index child only					
By gender					
Male	941	0.691		0.730	
Female	928	0.722	0.000	0.751	0.000
By wealth (Round 1)					
Bottom quintile	371	0.723		0.752	
Top quintile	374	0.685	0.000	0.722	0.000
By urban–rural					
Urban	1,374	0.699		0.735	
Rural	494	0.728	0.000	0.756	0.000
By maternal schooling attainment					
Less than primary	646	0.722		0.753	
Complete primary	527	0.709	0.015	0.743	0.007
Complete secondary	376	0.697	0.000	0.734	0.000
Higher education	255	0.676	0.000	0.716	0.000
By maternal native tongue					
Non-Spanish	524	0.716		0.747	
Spanish	1,314	0.703	0.006	0.738	0.013

Ethiopia	n	Outcome	p-value	Baseline	p-value
Index children and younger siblings					
Average					
Index child	1,810	0.719		0.725	
Younger sibling	937	0.764	0.000	0.772	0.000
By age					
7 or less	124	0.780		0.777	
8	243	0.773	0.319	0.781	0.558
9	316	0.761	0.010	0.771	0.306
10	195	0.758	0.006	0.765	0.105
11	641	0.725	0.000	0.749	0.000
12	1,211	0.716	0.000	0.740	0.000
13	2	0.726	0.267	0.729	0.227
Index child only					
By gender					
Male	957	0.708		0.734	
Female	853	0.731	0.000	0.753	0.000
By wealth (Round 1)					
Bottom quintile	374	0.727		0.749	
Top quintile	354	0.693	0.000	0.722	0.000
By urban–rural					
Urban	669	0.703		0.732	
Rural	1,140	0.728	0.000	0.749	0.000
By maternal schooling attainment					
No schooling	878	0.726		0.748	
Lower primary (grades 1-4)	418	0.720	0.216	0.745	0.477
Upper primary (grades 5-8)	279	0.705	0.000	0.733	0.000
More than 8 grades	174	0.696	0.000	0.723	0.000
By region					
Addis Ababa	264	0.700		0.727	
Amhara	351	0.721	0.001	0.745	0.000
Oromia	387	0.732	0.000	0.752	0.000
SNNP	424	0.710	0.099	0.737	0.051
Tigray	380	0.727	0.000	0.749	0.000

Notes: For each child, trial-level data were collapsed into a unique observation (the average). P-values correspond to the results of t-tests testing the null hypothesis that the mean difference is equal to zero.

Appendix C

Table C1. *Data dictionary for long-term memory tasks*

Concept: Long-term memory (task 1 and 5)

Variable label	Variable name in STATA
Unique child identifier	childid
Identifies whether child is sibling or index child	sibling
RACER task number	task_num
Game round number	round
Trial number	trialnum
Equals 1 if correct at first touch	holeinone
Percentage of correct first touches, out of 18 trials	perc_holeinone_r2tor4
Percentage of correct first touches, out of six trials	perc_holdinone_r1
Hour of test, in increments of 4 hours	hr_by4
Equals 1 if test was done on a weekend	wkend
Number of practice trials	practices
Pair identifier	pair
Number of touches until correct pair	touches
Average number of touches until correct pair, per round	ave_touch_rnd
Number of correct first touches, per round	total_holeinone_rnd
Percentage of correct first touches, per round	perc_holeinone_rnd
Number of correct first touches, out of 18 trials	total_holeinone_r2tor4
Number of correct first touches, out of six trials	total_holeinone_r1
Date of test	mdy
Hour of test	hr
Day of the week	dow
Minutes between task 1 and task 6	timegap
Percentage of correct first touches out of 18 trials	flag

Table C2. *Data dictionary for inhibitory control task*

Concept: Inhibitory control (task 2)

Variable label	Variable name in STATA
Unique child identifier	childid
Identifies whether child is sibling or index child	sibling
RACER task number	task_num
Trial number for opposite side trials	trialnum_opp
Time to touch (seconds)	resptime_opp
Equals one if touched the correct side	correct_opp
Euclidean distance from touch to correct location, opposite side trials	euclideand_opp
Horizontal distance from touch to correct location, opposite side trials	horizontald_opp
Average reaction time for opposite side trials	resptimeave_opp
Average reaction time for same side trials	resptimeave_same
Average horizontal distance from touch to correct location, opposite side trials	horizontaldave_opp
Average Euclidean distance from touch to correct location, opposite side trials	euclideandave_opp
Average horizontal distance from touch to correct location, same side trials	horizontaldave_same
Average Euclidean distance from touch to correct location, same side trials	euclideandave_same
Number of same side time outs	timeoutn_same
Number of opposite side time outs	timeoutn_opp
Number of timeouts (out of all 60 trials)	timeoutn_all
Number of correct same side touches	correctn_same
Number of correct opposite side touches	correctn_opp
Number of correct touches (out of all 60 trials)	correctn_all
Hour of test, in increments of 4 hours	hr_by4
Equals 1 if test was done on a weekend	wkend
Number of practice trials	practices
Trial number, within task/block	trialnum
Timed out (no press within 2.5 seconds)	timeout
Date of test	mdy
Hour of test	hr
Day of the week	dow

Table C3. *Data dictionary for working memory task*

Concept: Working memory (task 3)

Variable label	Variable name in STATA
Unique child identifier	childid
Identifies whether child is sibling or index child	sibling
RACER task number	task_num
Trial type: number of dots	ldot
Euclidean distance from touch to the correct location	dev
Response time	time
Percentage of timeouts trials (out of all 42 trials)	timeoutn_all
Percentage of short delay trials valid and answered (out of all 42 trials)	delayn_all_0
Percentage of long delay trials valid and answered (out of all 42 trials)	delayn_all_1
Percentage of 1 dot trials valid and answered (out of all 42 trials)	dotn_all_0
Percentage of multiple dot trials valid and answered (out of all 42 trials)	dotn_all_1
Average Euclidean distance from touch to correct location, single dot long delay trials	euclidean_sdot_delay_1
Average Euclidean distance from touch to correct location, single dot long delay trials (log scale)	euclideandave_sdot_delay_1
Average Euclidean distance from touch to correct location, single dot short delay trials	euclidean_sdot_delay_0
Average Euclidean distance from touch to correct location, single dot short delay trials (log scale)	euclideandave_sdot_delay_0
Average Euclidean distance from touch to correct location, multiple dot short delay trials	euclidean_mdot_delay_0
Average Euclidean distance from touch to correct location, multiple dot short delay trials (log scale)	euclideandave_mdot_delay_0
Average Euclidean distance from touch to correct location, multiple dot trials	euclidean_dots_1
Average Euclidean distance from touch to correct location, multiple dot trials (log scale)	euclideandave_dots_1
Average Euclidean distance from touch to correct location, single dot trials	euclidean_dots_0
Average Euclidean distance from touch to correct location, single dot trials (log scale)	euclideandave_dots_0
Average Euclidean distance from touch to correct location, long delay trials	euclidean_delay_1
Average Euclidean distance from touch to correct location, long delay trials (log scale)	euclideandave_delay_1
Average Euclidean distance from touch to correct location, short delay trials	euclidean_delay_0
Average Euclidean distance from touch to correct location, short delay trials (log scale)	euclideandave_delay_0
Average Euclidean distance from touch to correct location, long delay single dot trials	euclidean_dot_0
Average Euclidean distance from touch to correct location, long delay single dot trials (log scale)	euclideandave_dot_0
Average Euclidean distance from touch to correct location, long delay multiple dot long delay trials	euclidean_dot_1
Average Euclidean distance from touch to correct location, long delay multiple dot long delay trials (log scale)	euclideandave_dot_1
Average response time, multiple dot trials	resptimeave_dot_1
Average response time, single dot trials	resptimeave_dot_0
Average response time, long delay trials	resptimeave_delay_1
Average response time, short delay trials	resptimeave_delay_0

Variable label	Variable name in STATA
Percentage of 2 or 3 dots trials valid and answered (out of all 42 trials)	dotn_all_1
Percentage of 1 dot trials valid and answered (out of all 42 trials)	dotn_all_0
Percentage of long delay trials valid and answered (out of all 42 trials)	delayn_all_1
Percentage of short delay trials valid and answered (out of all 42 trials)	delayn_all_0
Hour of test, in increments of 4 hours	hr_by4
Equals 1 if test was done on a weekend	wkend
Number of practice trials	practices
Trial number, within task/block	trialnum

Table C4. *Data dictionary for implicit learning task*

Concept: Implicit learning (task 4)

Variable label	Variable name in STATA
Unique child identifier	childid
Identifies whether child is sibling or index child	sibling
RACER task number	task_num
Block number	block
Trial number, within block	trialnum
Response time for trial (in seconds)	resptime_patt
Average response time of trials in blocks 3 and 5 (patterned)	resptimeave_patt
Average response time of trials in blocks 1 and 4 (unpatterned)	resptimeave_unpatt
Number of non-missing trials in blocks 3 and 5	count_patt
Number of non-missing trials in blocks 1 and 4	count_unpatt
Number of practice trials	practices
Hour of test, in increments of 4 hours	hr_by4
Equals 1 if test was done on a weekend	wkend
Date of test	mdy
Hour of test	hr
Day of the week	dow
1 if touched near dot; 0 if no touch or far from dot	correct
Average response time of trials in block 1	resptimeave_b1
Average response time of trials in block 2	resptimeave_b2
Average response time of trials in block 3	resptimeave_b3
Average response time of trials in block 4	resptimeave_b4
Average response time of trials in block 5	resptimeave_b5
Number of non-missing trials in block 1	count_b1
Number of non-missing trials in block 2	count_b2
Number of non-missing trials in block 3	count_b3
Number of non-missing trials in block 4	count_b4
Number of non-missing trials in block 5	count_b5

Young Lives

An International Study of Childhood Poverty

About Young Lives

Young Lives is an international study of childhood poverty and transitions to adulthood, following the lives of 12,000 children in four countries (Ethiopia, India, Peru and Vietnam). Young Lives is a collaborative research programme led by a team in the Department of International Development at the University of Oxford in association with research and policy partners in the four study countries.

Through researching different aspects of children's lives across time, we seek to improve policies and programmes for children and young people.

Young Lives Research and Policy Partners

Ethiopia

- *Policy Studies Institute*
- *Pankhurst Development Research and Consulting plc*

India (Andhra Pradesh and Telangana)

- *Centre for Economic and Social Studies, Hyderabad (CESS)*
- *Sri Padmavati Mahila Visvavidyalam (Women's University), Tirupati (SPMVV)*

Peru

- *Grupo de Análisis para el Desarollo (GRADE)*
- *Instituto de Investigación Nutricional (IIN)*

Vietnam

- *Centre for Analysis and Forecast, Viet Nam Academy of Social Sciences (CAF-VASS)*
- *General Statistics Office of Viet Nam (GSO)*

Contact:
Young Lives
Oxford Department of International Development,
University of Oxford,
3 Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@qeh.ox.ac.uk
Website: www.younglives.org.uk

UKaid from the British people

Young Lives