



## Research Paper

# The association of a novel digital tool for assessment of early childhood cognitive development, 'DEvelopmental assessment on an E-Platform (DEEP)', with growth in rural India: A proof of concept study

Supriya Bhavnani<sup>a,b</sup>, Debarati Mukherjee<sup>c</sup>, Sunil Bhopal<sup>d,e</sup>, Kamal Kant Sharma<sup>a</sup>, Jayashree Dasgupta<sup>a</sup>, Gauri Divan<sup>a</sup>, Seyi Soremekun<sup>f,g</sup>, Reetabrata Roy<sup>a,d</sup>, Betty Kirkwood<sup>d</sup>, Vikram Patel<sup>a,h,i,\*</sup>

<sup>a</sup> Child Development Group, Sangath, Goa, India

<sup>b</sup> Centre for Chronic Conditions and Injuries, Public Health Foundation of India, Gurgaon, India

<sup>c</sup> Indian Institute of Public Health-Hyderabad, Bengaluru Campus, Bengaluru, Karnataka, India

<sup>d</sup> Maternal & Child Health Intervention Research Group, Department of Population Health, Faculty of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

<sup>e</sup> Population Health Sciences Institute, Newcastle University, United Kingdom

<sup>f</sup> Department of Clinical Research, Faculty of Infectious and Tropical Disease, London School of Hygiene and Tropical Medicine, London United Kingdom

<sup>g</sup> Observational and Pragmatic Research Institute, Singapore

<sup>h</sup> Department of Global Health & Social Medicine, Harvard Medical School, United States

<sup>i</sup> Department of Global Health and Population, Harvard T H Chan School of Public Health, United States

## ARTICLE INFO

## Article History:

Received 8 February 2021

Revised 21 May 2021

Accepted 28 May 2021

Available online xxx

## Keywords:

Early childhood development

Serious game

Cognitive development

Digital assessment

## ABSTRACT

**Background:** There is an urgent need to fill the gap of scalable cognitive assessment tools for preschool children to enable identification of children at-risk of sub-optimal development and to support their timely referral into interventions. We present the associations between growth in early childhood, a well-established marker of cognitive development, and scores on a novel digital cognitive assessment tool called DEvelopmental Assessment on an E-Platform (DEEP) on a sample of 3-year old pre-schoolers from a rural region in north India.

**Methods:** Between February 2018 and March 2019, 1359 children from the Sustainable Programme Incorporating Nutrition and Games (SPRING) programme were followed up at 3-years age and data on DEEP, anthropometry and a clinical developmental assessment, the Bayley's Scale of Infant and Toddler Development, 3rd edition (BSID-III) was collected. DEEP data from 200 children was used to train a machine learning algorithm to predict their score on the cognitive domain of BSID-III. The DEEP score of the remaining 1159 children was then predicted using this algorithm to examine the cross-sectional and prospective association of growth with the DEEP score.

**Findings:** The magnitude of the concurrent positive association between height-for-age and cognitive z-scores in 3-year olds was similar when cognition was measured by BSID-III (0.20 standard deviations increase for every unit change in specifically age-adjusted height (HAZ), 95% CI = 0.06–0.35) and DEEP (0.26 CI, 0.11–0.41). A similar positive prospective relationship was found between growth at 18 (0.21 CI, 0.17–0.26) and 12-months (0.18 CI, 0.13–0.23) and DEEP score measured at 3-years. Additionally, the relationship between growth and cognitive development was found to be dependant on socioeconomic status (SES).

**Interpretation:** In this study, we suggest the utility of DEEP, a scalable, digital cognitive assessment tool, to measure cognition in preschool children. Further validation in different and larger datasets is necessary to confirm our findings.

**Funding:** The SPRING Programme was funded through a Wellcome Trust programme grant and the follow-up study by the Corporate Social Responsibility initiative grant from Madura Microfinance Ltd.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\* Corresponding author at: Harvard Medical School, 641 Huntington Ave, Boston, MA 02115, USA

E-mail address: [vikram\\_patel@hms.harvard.edu](mailto:vikram_patel@hms.harvard.edu) (V. Patel).

## Research in context

### *Evidence before this study*

There is an urgent need for scalable tools for the assessment of cognitive abilities in young children not only to enable identification of children who need interventions, but also to offer an approach for routine surveillance, similar to growth monitoring. The only methods currently available either require lengthy parental interviews or observations of children by skilled providers, the use of proprietary developmental assessment tools, or using proxy indicators related to growth.

### *Added value of this study*

This paper extends the proof-of-concept of a novel assessment of cognitive development for preschool children, the “DEvelopmental Assessment on an E-Platform” (DEEP) which comprises gamified age-appropriate neuropsychological tasks. A machine learning derived algorithm was generated on 200 children to predict a score on the cognitive domain of a clinical developmental assessment, the Bayley’s Scale of Infant and Toddler Development, 3rd edition (BSID-III). The resulting DEEP score was applied to a larger population sample of 1159 children and a concurrent positive association was found between height-for-age and DEEP scores in 3-year olds and a predictive association between growth at 12 and 18 months and the DEEP score at 3-years.

### *Implications of all the available evidence*

This study presents the degree to which DEEP, a tool which could potentially fill this gap of scalable cognitive developmental assessments for preschool children in global child health, is comparable with the BSID-III, a gold-standard cognitive assessment. Further validation in additional datasets is required to confirm these findings.

However, there are currently no scalable methods for such routine assessment of cognitive abilities in early childhood. Existing assessments rely on structured observations by highly trained specialists like developmental paediatricians or clinical psychologists, who are scarce and expensive resources in low and middle income countries (LAMIC). Further, the measures themselves, such as the Bayley’s Scale for Infant and Child Development (BSID), are not freely accessible, need adaptation for use in diverse contexts, and have a significant time burden for administration [8]. There is thus an urgent need to develop low-cost scalable tools for assessment of cognitive development that can be used by non-specialists in diverse settings [9]. The development, validation and deployment of tools like Guide for Monitoring Child Development (GMCD) and Early Child Development Index (ECDI) which can be administered by non-specialists represents efforts in this direction [10,11]. However, these parent-report questionnaires rely on parents’ (a) *knowledge* of age-appropriate cognitive milestones in early childhood, (b) abilities to closely observe their own children’s behaviours and *recognise* faltering development, and (c) willingness to acknowledge and *report* missed milestones; all of which potentially contribute to lower sensitivity, especially in households living in poverty. Given the emerging evidence of mobile devices like tablet computers and smartphones penetrating health systems of many countries [12], tools that harness these technologies to directly assess the child present potential solutions to these challenges [13].

In order to address the gap of scalable cognitive assessment tools which directly measure child performance, our team has developed a digital tool called “DEvelopmental Assessment on an E-Platform” (DEEP) which comprises gamified age-appropriate neuropsychological tasks for preschool children [14]. These tasks have been designed to be culturally-agnostic through the use of universally relatable images, and are woven into a first-person narrative story with the moon and a child as protagonists. The DEEP games (see eFig.1 for game snapshots and brief descriptions) assess multiple cognitive skills including manual processing speed and coordination, attention, response inhibition, reasoning, visual form perception and integration, and memory. DEEP has been piloted on a cohort of children in a rural north Indian region and demonstrated to be highly engaging for children across genders, acceptable to their parents and feasible for delivery by trained non-specialist personnel in the comfort of the child’s home [14]. A proof-of-concept study has also demonstrated that it is possible to predict children’s score on the cognitive domain of the Bayley’s Scale for Infant and Toddler Development (BSID – IIIrd edition), using metrics captured by DEEP [15]. This study used a supervised machine learning approach benchmarked to the BSID-III cognitive score to develop an algorithm comprising a combination of features extracted from a child’s performance on different DEEP games to derive the DEEP cognitive score.

The study presented in this paper aims to extend the proof-of-concept of DEEP’s utility as a cognitive assessment tool for preschool children. Firstly, we compare the distribution of BSID-III cognitive domain and DEEP scores in a sample of 3-year old children ( $N = 200$ ) from a rural region in north India. Secondly, we examine the association between DEEP score and growth measures in early childhood, specifically age-adjusted height (HAZ), which is a well-established marker of early childhood cognitive development [16–19]. Poor physical growth in utero and until 3-years of age, as a result of exposure to infections and chronic poor nutrition, results in stunting which is defined as HAZ being two standard deviations below the WHO median values. HAZ has been consistently demonstrated to be positively associated with academic performance, with non-stunted children having more years in education and higher income potential [20]. It has thus been commonly used as a proxy indicator for generating global and regional estimates of children at-risk for not developing optimally [20,21]. To this end, we examine: (1) the cross-sectional association between HAZ and cognitive development as

## 1. Introduction

The preschool years of childhood, from birth to 6-years age, represents a critical developmental period when the brain’s structural and functional development rate is at its peak [1]. During this period, children acquire crucial cognitive skills which include attention, inhibitory control, visuo-motor coordination and memory, that allow them to process information [2]. These developmental processes are sensitive to a range of risk factors, in particular those associated with poverty, leading to delays in cognitive development with adverse health and economic consequences across the life course. These risk factors include inadequate nutrition and exposure to infectious diseases, absence of an enriched environment providing cognitive stimulation and presence of maternal stress and depression [3,4]. The preschool years are also the time at which the brain is most plastic and amenable to change and thereby responsive to effective interventions [5]. The importance of investing in this period of early childhood is globally recognised, as demonstrated through the inclusion of Sustainable Development Goal (SDG) 4 in the framework of the SDGs [6] and the Nurturing Care Framework recently published by the World Health Organisation (WHO) and United Nations Children’s Fund (UNICEF) [7]. An essential step to ensuring that all children have the opportunity to thrive and reach their full developmental potential is the regular monitoring of cognitive abilities to enable early identification of those that are faltering in their development so as to support their timely referral to interventions which would be most effective during this age [5].

measured by BSID-III on a sub-sample of this population of 3-year olds ( $N = 200$ ), and compare it to associations with DEEP on the same population ( $N = 200$ ) and whole sample ( $N = 1356$ ); and 2) the prospective association between HAZ at 12 and 18 months of age, and cognitive development as measured by DEEP and BSID-III at 3 years. Thirdly, we examined the extent to which DEEP adds to the prediction of the BSID cognitive domain score from HAZ alone.

## 2. Methods

### 2.1. Study design and participants

Participants in this study were recruited from 120 villages in Rewari district in rural Haryana, India, the site for the SPRING (Sustainable Programme Incorporating Nutrition and Games) trial, an early childhood development randomised control trial, which has been described in detail elsewhere [22,23]. In brief, SPRING developed an innovative, feasible, affordable & sustainable community-based approach to delivering a home visiting programme aiming to improve child growth & development at-scale in India & Pakistan (registered with Clinical-Trials.gov, number NCT02059863). 7015 children were enrolled into SPRING's surveillance system, with 5117 born from 18 June 2015 when the SPRING intervention was fully implemented and therefore eligible for recruitment into the trial. Of these 1744 were identified for the child development assessment sub-sample with the aim of assessing at least 50 children in each cluster at age 18 months. The loss to follow-up was less than expected and 1443 children therefore received an anthropometric assessment at 18-months age.

Between February and May 2018, 100 3-year old children from SPRING's surveillance system were randomly selected and assessed as part of a DEEP pilot study which collected DEEP, BSID-III and growth data. Subsequently, between August 2018 and March 2019, an additional 1259 of the 1443 children who received an assessment of anthropometry at 18-months age, were assessed when they were approximately 3-years old. DEEP and growth data was collected on all 1259, while BSID-III data was collected on a subset of 100 of these children. 184 children were lost to follow up due to the following reasons: 122 had moved away from the study area, 40 were temporarily unavailable during the assessment period, 12 families refused consent, 2 children were unable to engage with the tablet due to a physical disability and 3 had died. DEEP data did not save for another 5 children. The entire sample of this study is constituted by 1359 children (study samples summarised in Table 1). A comparison of the socio-demographic profile of study children with the remaining children enrolled in the SPRING surveillance system ( $N = 5656$ ) can be found in eTable 1.

All assessments on 3-year old children followed up through this study were conducted by eight non-specialists (henceforth referred to as 'assessors') in participants' households at a convenient date and time. These assessors had completed the equivalent of a post-graduate degree, and had been part of the SPRING evaluation teams. They were thus embedded within the community and had prior training and experience working with young children. Approximately 10% of all visits were supervised by a field supervisor. Weekly group

meetings between the field supervisor and all assessors were used to provide peer support and regular feedback, and quarterly refresher trainings were conducted by senior research team members.

Written informed consent was taken from parents at enrolment in SPRING, then prior to the 18-month assessment. Parents were also consented to be approached by the research team after completion of the SPRING trial. Written informed consent was again obtained at the time of the 3-year follow-up assessment. Ethics approval for SPRING was obtained from the London School of Hygiene & Tropical Medicine (LSHTM) research ethics committee (23 June 2011; approval number 5983) and the Sangath Institutional Review board (IRB) (19 February 2014). Approval was also granted by the Indian Council of Medical Research's Health Ministry Screening Committee (HMSC) (24 November 2014). Ethical approval for the study which collected the data reported in this paper was obtained from IRBs of Public Health Foundation of India (PHFI) (27 October 2017; 18 July 2018), Sangath (23 August 2018), and the LSHTM research ethics committee (11 June 2020; approval number 9886 (5983) – 6).

### 2.2. Data collection and preparation

**Cognition at 3-years of age:** The Bayley's Scale of Infant and Toddler Development, 3rd Edition (BSID-III), a developmental assessment for preschool children aged 0–42 months [24], was administered on the 200 participants described above. A translated version of the BSID-III adapted for administration by non-specialists was used following a protocol described previously [15,22]. Raw scores were computed as per the manual, and used to generate age-adjusted composite scores.

DEEP (see eFig.1) was administered on Samsung Tab E Android tablets. At the beginning of each of DEEP's 9 games, assessors delivered standardised verbal instructions in the local language most familiar to the child to teach them how to play the games (demo-mode) [14]. To ensure that comprehension of language was not a limiting factor in the child's ability to understand the instructions, the assessor would first show the child how to play the game, and then assist the child till they were able to play independently. Assessors were trained to proceed to play-mode only when a child could engage independently and correctly with the demo-mode without any assistance.

**Anthropometry:** World Health Organisation (WHO) protocols were used to measure the child's height using the SECA-417 infantometer at 12 and 18 months and Seca 213 Portable Stadiometer and 3 years. Height was used to generate height-for-age (HAZ) z-scores using WHO growth standards. Stunting was defined as two standard deviations below the age-adjusted WHO growth-standard median values of height. All children whose age-adjusted anthropometric measurements were below three standard deviations of WHO median values were referred for follow-up assessments to local clinics.

**Socioeconomic status:** Information on socioeconomic status was collected from families upon enrolment into the SPRING study [22]. Principal components analysis was used to calculate a socioeconomic status (SES) index using data on household demographics and animal & other asset ownership. This index was used to categorize the population into SES quintiles.

**Table 1.**  
Sample details (source, N number) and types of analyses in this study. HAZ= height-for-age.

Age at measurement of predictor	Analysis type	DEEP score			BSID-III cognitive domain score		
		SPRING surveillance arm	SPRING outcome arm	Total	SPRING surveillance arm	SPRING outcome arm	Total
3-year HAZ	(A) Concurrent association	100	1259	1359	100	100	200
18-month HAZ	(B) Prospective association	–	1259	1259	–	100	100
12-month HAZ	(B) Prospective association	–	1122	1122	–	70	70

### 2.3. Statistical analysis

#### 2.3.1. Predicting BSID-III cognitive score from DEEP features

Child performance during the play-mode of DEEP's games were the features on which a supervised machine learning approach was applied to derive the optimal combination of features for the prediction of BSID-III cognitive domain scores. Briefly, this approach utilised seven feature selection methods in combination with five prediction functions to identify the best model using a 10-fold cross-validation procedure and ensemble modelling repeated 10 times for stability (refer to supplementary material for details). The metrics used to assess the performance of this model were: (a) Pearson's correlation coefficient between the predicted (DEEP score) and true BSID-III cognitive domain raw score, (b) absolute agreement, using two-way, random effects intra-class coefficient [ICC(2,1)], (c) mean absolute prediction error defined as DEEP score – BSID-III cognitive domain score and d) root mean square error. This algorithm was trained using the full sample of children with both BSID-III and DEEP data available at 3-years ( $N = 200$ ), and then used to generate the DEEP score for the remaining 1159 children. The distribution of these scores was tested for normality using the Shapiro Wilk W Test and compared using a two-way Kolmogorov–Smirnov test.

#### 2.3.2. Associations between growth and cognitive development

Three types of associations were conducted in this study (see [Table 1](#)): (a) cross-sectional associations between HAZ and cognitive development, assessed using BSID-III and DEEP, measured concurrently at 3-years and (b) prospective associations between 12 and 18-month HAZ with cognitive development measured using BSID-III and DEEP at 3-years. To allow for comparison of results across the two measures of cognition, BSID-III and DEEP scores were converted to standard z-scores. All associations were tested using mixed effects linear regression (xtmixed function) on Stata version 14, with the trial-cluster in which the child resides being used as a random-effect and predictor variables being treated as fixed-effects. Sex and SES were explored for interaction with HAZ; the interaction term for SES, but not for sex, was found to be significant. Thus sex was included in the first model for each analysis (Model 1) as a confounder, along with child age at visit and SPRING trial arm allocation. In addition to these, SES was included as an interaction term in the second model (Model 2). For each regression model the mean cognitive z-score (BSID-III or DEEP) at mean HAZ, and up to 2 standard deviations above and below the mean HAZ (i.e. in order to assess the spread of the predicted score), was predicted for every SES quintile level and plotted using Microsoft Excel. Finally, a hierarchical regression was conducted in which the Likelihood-Ratio Test was used to compare the goodness of fit of a model in which BSID-III was predicted using 3-year HAZ alone with a model which also had DEEP score as a predictor variable. The null hypothesis is that the smaller model is the "best" model which is rejected when the test statistic is large.

#### 2.3.3. Role of the funding source

The funding agencies had no involvement in the collection, analysis, and interpretation of data, writing of the report and the decision to submit the paper for publication. All authors had full access to the full data in the study and had final responsibility for the decision to submit for publication.

## 3. Results

### 3.1. Description of study participants

The socio-demographic and growth profile of the children followed up at 3-years age is presented in [Table 2](#). The mean age of children assessed in the follow-up study was 38.7 months (standard deviation (SD) = 1.1 months), just under half (45.8%) were girls. While

**Table 2.**

Socio-demographic and growth profile of study participants at 3-years age. Parental education and SES data were collected at enrolment into SPRING.

Characteristic	N = 1359
Female, n (%)	623 (45.9)
Age (months), mean (sd)	38.7 (1.1)
Mother's age at delivery, mean (sd)	22.3 (3.8)
Mother's education level, n (%)	
Below primary (including never been to school)	168 (12.4)
Primary/middle school completed	350 (25.8)
Secondary/higher secondary school completed	525 (38.6)
College & above	316 (23.3)
Father's education level, n (%)	
Below primary (including never been to school)	72 (5.3)
Primary/middle school completed	268 (19.7)
Secondary/higher secondary school completed	613 (45.1)
College & above	406 (29.9)
SES quintile, n (%)	
Q1 (poorest)	282 (20.8)
Q2	306 (22.5)
Q3	273 (20.1)
Q4	264 (19.4)
Q5 (wealthiest)	234 (17.2)
Height-for-age (z-score), mean (95% CI) <sup>#</sup>	-1.58 (-3.5–0.4)
Stunted, n (%)	439 (32.4)
Preschool enrolment, n (%)	
Private preschool	330 (24.3)
Anganwadi centres	261 (19.2)
None	768 (56.5)
BSID-III cognitive domain score <sup>§</sup>	
Range	57–88
Mean (SD)	69.41 (5.02)
Median (IQR)	69 (7)
DEEP score <sup>§</sup>	
Range	61.72–76.06
Mean (SD)	69.40 (3.26)
Median (IQR)	69.23 (4.93)
DEEP score	
Range	60.45–79.25
Mean (SD)	69.76 (3.13)
Median (IQR)	69.87 (4.84)

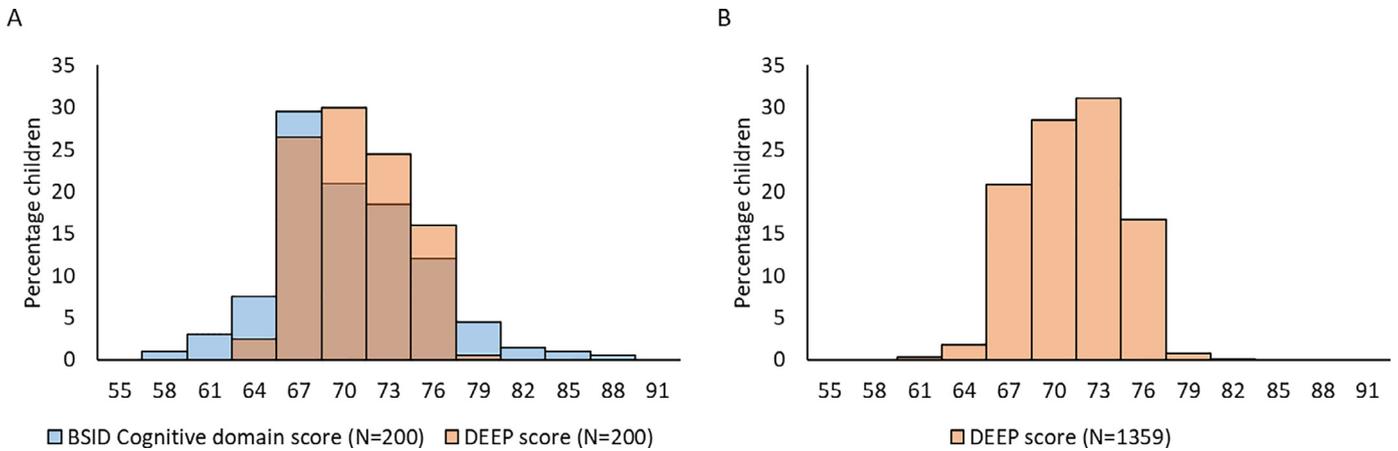
<sup>#</sup> n = 1356 <sup>§</sup> n=subsample of 200.

the greater proportion of both mothers and fathers were educated till secondary or higher secondary school, more fathers than mothers were educated up to this level. Consistent with prior reports from this population, we observed that 32.4% of the children were stunted. More than half the children (56.5%) were not attending any preschool at 3-years age. Participating families were almost equally distributed across SES quintiles, with slightly more (22.5%) and fewer (17.2%) children in quintiles 2 and 5, respectively. The study sample thus had a slight overrepresentation of children from the poorer quintiles, as seen in [eTable 1](#) which compares it with children enrolled in the SPRING surveillance system but not followed up at 3-years.

### 3.2. Predicting BSID-III cognitive score from DEEP features

Of the 1359 children on whom DEEP was administered, we found that 1342 (98.7%) attempted all nine games. The time taken to complete the DEEP assessment depends on the number of difficulty levels attempted by each child; we found that on average children spent just over 20 min (mean = 23 min 18 s; SD = 3 min 52 s) engaged with the tool. This duration ranged from under 3 min in children who only played the first couple of games till over 50 min for those who successfully completed most game levels (range = 3 min 39 s–50 min 40 s).

The distribution of both the predicted (DEEP score) and true BSID-III cognitive domain scores were found to be non-normal (BSID-III  $N = 200$   $z = 3.05$ ,  $p = 0.001$ ; DEEP  $N = 200$   $z = 2.63$ ,  $p = 0.004$ ; DEEP  $N = 1359$   $z = 6.17$ ,  $p < 0.001$ ) and unequal as determined by the KS Test comparing BSID-III and DEEP on the sample of 200 children



**Fig. 1.** Distribution of scores of cognitive assessments. Distribution of (A) the BSID cognitive score (blue) and DEEP score (orange) of 200 children on whom the machine learning algorithm was trained and (B) the DEEP score of the full sample of 1359 children for whom it was derived (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

( $p = 0.02$ ) (Fig. 1A). The BSID-III cognitive domain score of 200 children ranged from 57 to 88 (mean = 69.41, SD = 5.02; median = 69, IQR = 7), while the DEEP score of the same children had a smaller range of 61.72 to 76.06 (mean = 69.40, SD = 3.26; median = 69.23, inter-quartile range (IQR) = 4.93) (see Table 2). We found that the performance, including limitations, of the machine learning algorithm derived in this study was comparable to the previously published algorithm (eTable2). For instance, while the overall correlation between the DEEP score and BSID-III cognitive domain score (Pearson’s correlation coefficient  $r = 0.673$ ) (eFig.2A) and absolute agreement were moderate (ICC (2,1) = 0.62, 95% CI = 0.52–0.70), DEEP tended to overestimate low BSID scorers and underestimate high scorers (eFig.2B). The mean absolute prediction error of the ML algorithm was 2.91 (SD = 2.31) and root mean square error was 3.71 (SD = 4.71). The DEEP score of the entire population (Fig. 1B), predicted using this algorithm, ranged from 60.45 to 79.25 (mean = 69.76, SD = 3.13; median = 69.87, IQR = 4.84).

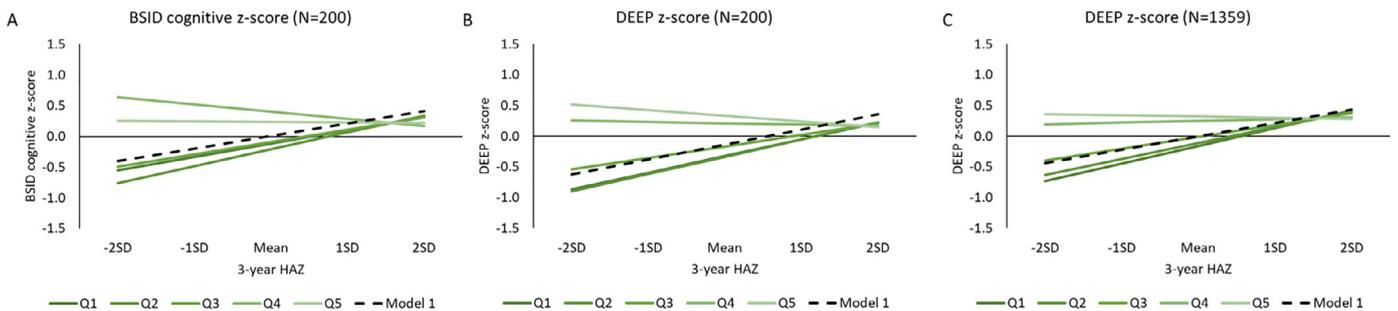
**3.3. Associations between HAZ and cognitive development measured concurrently at 3-years age**

A positive relationship was observed between HAZ and cognitive development as measured by BSID-III ( $N = 200$ ) and DEEP ( $N = 200$  and  $N = 1356$ ) (see eFig.3); for every unit change in HAZ, the BSID cognitive z-score increased by 0.20 standard deviations (95% CI = 0.06–0.35), comparable to the 0.26 (0.11–0.41)

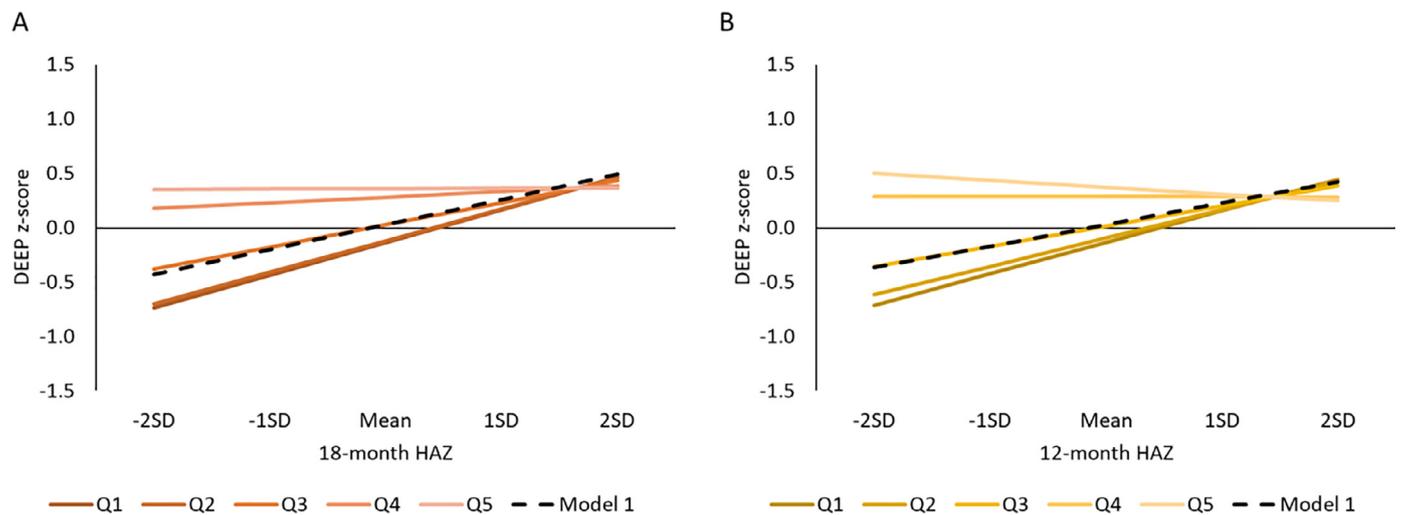
SD unit increase in DEEP z-score for the same population (Model 1 slope represented as black dotted line in Fig. 2A and B, respectively, see eTable3). SES modified this relationship with the slope of the regression line highest in the poorest quintile and decreasing as the SES quintile increased. Thus, while there was a strong positive association between HAZ and cognitive development in low SES quintiles (1, 2 and 3), this relationship was not present in the wealthier SES quintiles (4 and 5) (see Fig. 2 and eTable3). This effect was observed irrespective of whether cognition was measured using BSID-III ( $N = 200$ ) or DEEP ( $N = 200$  and  $N = 1356$ ). Additionally, using hierarchical regression, we found that including DEEP as a predictor in associations between 3-year HAZ and BSID-III cognitive domain score significantly improved the model (LR  $\chi^2(1) = 69.71, p < 0.001$ ).

**3.4. Predictive associations between HAZ 12 and 18 months and cognitive development at 3-years age**

The relationship between HAZ, measured in the first 2 years of life, and cognitive development measured by DEEP and BSID-III at 3-years age, was similar to that with concurrently measured HAZ (see eTables 4 and 5). Children who were tallest at 12- and 18-months age had the highest cognitive z-scores at 3-years (DEEP Model 1 slope represented as black dotted line in Fig. 3A and B). This association was also modified by SES, such that there was a strong positive association in low SES quintiles (1, 2 and 3) but none in the wealthier SES quintiles (4 and 5) (see Fig. 3 and eTables 4 and 5).



**Fig. 2.** Associations between height-for-age (HAZ) and cognitive development concurrently measured at 3 years of age. Association between HAZ and (A) BSID-III cognitive z-score ( $N = 200$ ); (B) DEEP z-score ( $N = 200$ ); (C) DEEP z-score ( $N = 1359$ ) stratified by socioeconomic status quintile with Q1 (darkest green) and Q5 (lightest green) being the lowest and highest quintiles, respectively. Unadjusted models are represented as dashed lines (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Prospective associations between height-for-age (HAZ) and cognitive development. Associations between DEEP z-scores of 3-year old children and HAZ measured at (A) 18-months and (B) 12-months of age stratified by socioeconomic status quintile with Q1 (darkest) and Q5 (lightest) being the lowest and highest quintiles, respectively. Unadjusted models are represented as dashed lines (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

#### 4. Discussion

In this study, we used the well-documented relationship between physical growth and cognitive development to provide a proof-of-concept of the potential utility of DEvelopmental Assessment on an E-Platform (DEEP), a scalable tablet-based neurodevelopmental assessment tool, to index cognition in preschool children. We also reveal a complex relationship between growth and cognitive development, which is modified by the family's socioeconomic status. We discuss these results in the context of the need to fill the gap of scalable cognitive developmental assessments for preschool children.

In this study non-specialists administered DEEP to assess cognitive development on over 1350 young children in rural households. The socio-demographic profile of the children who participated in this study was found to be comparable to the SPRING study sample from which they were drawn, with a slight over-representation of the poorest socio-economic status quintiles. Also, importantly, key indicators of rate of stunting in children under 5 years and percentage of women with 10 or more years of schooling in population-wide surveys conducted in this district [25] are comparable to our study sample, suggesting that it can be considered to be representative of this region.

In this study we take our previous proof of concept of using a machine learning approach to predict children's BSID-III cognitive domain score through their performance on DEEP [15] a step further by refining our algorithm by using a slightly larger (200 children compared to 140 in the previous study) training dataset, albeit with the limitation of not having an independent validation dataset.

In this study population, using this prediction algorithm, we demonstrate a positive cross-sectional association between HAZ and cognitive development measured by BSID-III, a widely used clinical developmental assessment replicating the findings of other studies [18,19,26,27]. Importantly, we also observed that the coefficient of the concurrent association between HAZ and cognitive z-scores is similar when cognition is measured by BSID-III and by our novel DEEP assessment. We also demonstrate a positive prospective relationship between HAZ at 12 and 18-months-age and DEEP score at 3-years. These results provide evidence of the comparability of the performance of DEEP with the BSID-III as a cognitive assessment tool [3,28]. However, we acknowledge that these results are based on a machine learning algorithm which has not been validated on an independent dataset (see 'Strengths and limitations of the study' section of Discussion).

Our results indicate that while there is a positive cross-sectional and prospective relationship between height in early childhood and cognition at 3-years, measured either by the clinical assessment BSID-III or DEEP, this relationship is more pronounced in the most deprived socio-economic groups. Other studies have also reported such differential effects of family SES. Of particular relevance are two studies which have used the caregiver-report tool, Early Child Development Index (ECDI), to evaluate cognitive or socioemotional development in early childhood and examine associations with physical growth. Both studies have demonstrated that the association between physical growth and child development a) reduced significantly when adjusting for household SES and b) is moderate in countries with low human development index (HDI), with no association in high HDI countries [29,30].

These findings suggest that while both undernourishment and poor cognitive development in early childhood are caused by poverty, likely due to risk factors for both commonly co-occurring in poor households [3,4], the association between the two is moderated by other factors. We speculate that the most likely alternative factor is an enriched environment with greater availability of play materials and more responsive interactions between caregivers and children, which may be more common in households with higher SES, and which might be able to compensate for the impact of chronic under-nutrition impacting brain growth. The central role of a stimulating home environment on the development of children's cognitive abilities [5,31] is supported by numerous randomised control trials which have shown that psychosocial interventions targeting responsive parenting and cognitive stimulation have a larger impact on cognitive outcomes than nutritional interventions, and outcomes are enhanced when these two components are provided together [32,33].

The complexity of the relationship between anthropometry and cognitive development serves as a reminder of the need to use specific measures to assess children's developmental status rather than depending on proxies based on growth outcomes [27,29,30]. Indeed, authors of studies using stunting as a proxy indicator of cognitive development themselves suggest that this might be underestimating the true magnitude of the number of children not being able to develop optimally [20,21]. In support of this speculation, the results in this study also demonstrate that adding DEEP score as a predictor improves the association between HAZ and cognition. Realistic estimates of children at risk of not being able to attain their full developmental potential can thus only be achieved if the barrier of lack of availability of effective and scalable tools for cognitive assessments in

preschool children is overcome [9]. This study presents the degree to which DEEP, a tool which could potentially fill this gap in global child health, is comparable with the BSID-III, a gold-standard cognitive assessment. We demonstrate that even though in its current version which 3-year old children require almost 25 min to complete, DEEP is highly engaging to them as almost all children attempted every game. A high level of acceptability to the end-user, in this case preschool children who are notoriously difficult to engage in assessments, is an integral barrier to overcome when developing tools intended to be used at scale. We are now testing DEEP with larger samples in diverse populations to further evaluate its convergent, construct and cross-cultural validity, along with its test-retest reliability of DEEP, in accordance with the COSMIN guidelines [34] (<https://research.reading.ac.uk/stream/>). We also intend to use this data to make the administration of DEEP age-contingent, so as to reduce the burden of the assessment without compromising its validity.

The large study sample being community-based and representative of the population from which they were drawn serve as strengths of this study. Further, being a longitudinal follow-up of a birth cohort, it has allowed us to examine both cross-sectional and prospective associations between growth and cognitive development. A methodological limitation of this study is that DEEP score has been predicted from an ML algorithm has been trained on a relatively small sample size of 200 children. Further, unlike in our previous study which had split this sample into a training and test dataset, this time we have used the entire sample to train the model and do not present an independent test dataset. Our predictions are overestimating and underestimating poor and good BSID-III performers respectively which has resulted in them having unequal distributions of their cognitive scores. Despite these limitations of a small sample size for the training set and lack of an independent test set, our approach to using ML to analyse data from gamified cognitive assessments remains novel and our ongoing endeavour to expand the size and diversity of our sample, alongside gold-standard assessments, will allow us to derive more accurate and generalizable algorithms. Other study limitations include: (1) a small proportion of the SPRING sample (12.3%) was lost to follow-up, mostly due to the family moving to another location which was outside of the study area, (2) some of the children followed up in this study had missing data in their 12-month physical growth measures, and (3) household socio-economic status was calculated at enrolment to SPRING, and this may have changed prior to the 3-year assessment. Despite these limitations, to our knowledge, this is the first published study comparing the performance of a scalable, digital assessment tool which has the potential to contribute to filling the gap of global data on cognitive development in preschool years, with a gold-standard cognitive assessment tool through associations with an established proxy measure of cognition.

### Funding

These studies were funded through two awards: a Wellcome Trust programme grant for the SPRING Programme and a Corporate Social Responsibility (CSR) initiative grant from Madura Microfinance Ltd for the follow-up study.

### Data sharing statement

The datasets generated for this study are available on request to the corresponding author.

### Author contributions

SB, DM, S Bhopal, GD, VP and BK conceptualised the study. KKS coordinated and supervised all data collection while JD supervised BSID administration. SB conducted the formal analysis, supported by

DM, S Bhopal, RR and SS, visualised the data and prepared the original draft. All co-authors reviewed and edited the manuscript and agreed upon its final version.

### Declaration of Competing Interest

Dr Sharma reports grants from Madura Microfinance Ltd and Wellcome Trusts, during the conduct of the study. Dr. Kirkwood reports grants from Wellcome Trust, during the conduct of the study; Dr. Roy reports grants from Madura Microfinance Ltd, grants from Wellcome Trust, during the conduct of the study; Dr. Soremekun reports grants from Wellcome Trust, during the conduct of the study; Dr. Bhopal reports grants from Wellcome Trust, during the conduct of the study; Dr. Bhavnani reports grants from Madura Microfinance Ltd, grants from Wellcome Trust, during the conduct of the study; Dr. Patel reports grants from Madura Microfinance CSR, grants from Wellcome Trust, during the conduct of the study. All the other authors report non conflicts.

### Acknowledgments

We would like to acknowledge the contributions of Ms Deepali Verma and Ms Divya Kumar to the collection of data on the SPRING study and its follow-up contributing to this manuscript, and Dr Sidhartha Mandal for advice on the statistical analysis methods.

### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.eclinm.2021.100964](https://doi.org/10.1016/j.eclinm.2021.100964).

### References

- [1] Estrin GL, Bhavnani S. Brain development: structure editor. In: Benson JB, editor. *Encyclopedia of infant and early childhood development*. (2nd Ed.) Oxford: Elsevier; 2020. p. 205–14.
- [2] Johnson MH, de Haan M. *Developmental cognitive neuroscience: an introduction*. 4th Ed. Wiley-Blackwell; 2015. p. 332..
- [3] Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, et al. Early childhood development coming of age: science through the life course. *Lancet* 2017;389:77–90.
- [4] Nelson CA, Gabard-Durnam LJ. Early adversity and critical periods: neurodevelopmental consequences of violating the expectable environment. *Trends Neurosci* 2020;43:133–43.
- [5] Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, et al. Nurturing care: promoting early childhood development. *The Lancet* 2017;389:91–102.
- [6] Sustainable Development Goals [Internet]. UNDP. [cited 2020 Dec 21]. Available from: <https://www.undp.org/content/undp/en/home/sustainable-development-goals.html>
- [7] UNICEF, World Bank, World Health Organization. *Nurturing care for early childhood development a framework for helping children survive and thrive to transform health and human potential*. Geneva: World Health Organization; 2018. [Internet][cited 2019 Nov 15]. Available from: [https://www.who.int/maternal-child\\_adolescent/documents/nurturing-care-early-childhood-development/en](https://www.who.int/maternal-child_adolescent/documents/nurturing-care-early-childhood-development/en).
- [8] Marlow M, Servili C, Tomlinson M. A review of screening tools for the identification of autism spectrum disorders and developmental delay in infants and young children: recommendations for use in low- and middle-income countries. *Autism Res* 2019;12:176–99.
- [9] Boggs D, Milner KM, Chandna J, Black M, Cavallera V, Dua T, et al. Rating early child development outcome measurement tools for routine health programme use. *Arch Dis Child* 2019;104:S22–33.
- [10] Ertem IO, Dogan DG, Gok CG, Kizilates SU, Caliskan A, Atay G, et al. A guide for monitoring child development in low- and middle-income countries. *Pediatrics* 2008;121:e581–9.
- [11] Loizillon A, Petrowski N, Britto P, Cappa C. Development of the early childhood development index in MICS surveys. UNICEF; 2017 MICS Methodological Papers, No. 6, Data and Analytics Section, Division of DataResearch and Policy.
- [12] eHealth wgo for. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization; 2011. [Internet] [cited 2021 Apr 1]. Available from: <https://apps.who.int/iris/handle/10665/44607>.
- [13] McHenry M.S., Mukherjee D., Bhavnani S., Kirolos A., Piper J.D., Crespo-Llado M., et al. The future of tablet-based assessments to measure for the evaluation of pediatric cognition in low resourced settings. Submitted.
- [14] Bhavnani S, Mukherjee D, Dasgupta J, Verma D, Parameshwaran D, Divan G, et al. Development, feasibility and acceptability of a gamified cognitive Developmental

- assessment on an E-Platform (DEEP) in rural Indian pre-schoolers – a pilot study. *Glob Health Action* 2019;12:1548005. doi: 10.1080/16549716.2018.1548005.
- [15] Mukherjee D, Bhavnani S, Swaminathan A, Verma D, Parameshwaran D, Divan G, et al. Proof of concept of a gamified developmental assessment on an E-platform (DEEP) tool to measure cognitive development in rural Indian preschool children. *Front Psychol* 2020;11:1202.
- [16] Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, et al. Inequality in early childhood: risk and protective factors for early child development. *The Lancet* 2011;378:1325–38.
- [17] Crookston BT, Schott W, Cueto S, Dearden KA, Engle P, Georgiadis A, et al. Postinfancy growth, schooling, and cognitive achievement: young Lives. *Am J Clin Nutr* 2013;98:1555–63.
- [18] Miller AC, Murray MB, Thomson DR, Arbour MC. How consistent are associations between stunting and child development? Evidence from a meta-analysis of associations between stunting and multidimensional child development in fifteen low- and middle-income countries. *Public Health Nutr* 2016;19:1339–47.
- [19] Sudfeld CR, McCoy DC, Danaei G, Fink G, Ezzati M, Andrews KG, et al. Linear growth and child development in low- and middle-income countries: a meta-analysis. *Pediatrics* 2015;135:e1266–75.
- [20] Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B, et al. Developmental potential in the first 5 years for children in developing countries. *Lancet* 2007;369:60–70.
- [21] Lu C, Black MM, Richter LM. Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level. *Lancet Glob Health* 2016;4:e916–22.
- [22] Bhopal S, Roy R, Verma D, Kumar D, Avan B, Khan B, et al. Impact of adversity on early childhood growth & development in rural India: findings from the early life stress sub-study of the SPRING cluster randomised controlled trial (SPRING-ELS). *PLoS ONE* 2019;14:e0209122. doi: 10.1371/journal.pone.0209122.
- [23] London School of Hygiene and Tropical Medicine. Spring (for the millennium development goals): sustainable programme incorporating nutrition and games (for maximising development, growth and survival) [Internet]. [clinicaltrials.gov/ct2/show/NCT02059863](https://clinicaltrials.gov/ct2/show/NCT02059863). 2017 Aug [cited 2020 Dec 19]. Report No.: NCT02059863. Available from: <https://clinicaltrials.gov/ct2/show/NCT02059863>
- [24] Bayley Bayley N. Scales of infant and toddler development—third edition. Pearson Clinical; 2006.
- [25] National family health survey [internet]. [cited 2020 Dec 21]. Available from: <http://rchiips.org/nfhs/HR.shtml>
- [26] Sudfeld CR, McCoy DC, Fink G, Muhimi A, Bellinger DC, Masanja H, et al. Malnutrition and its determinants are associated with suboptimal cognitive, communication, and motor development in Tanzanian children. *J Nutr* 2015;145:2705–14.
- [27] Scharf RJ, Rogawski ET, Murray-Kolb LE, Maphula A, Svensen E, Tofail F, et al. Early childhood growth and cognitive outcomes: findings from the MAL-ED study. *Matern Child Nutr* 2018;14:e12584.
- [28] Kang Y, Aguayo VM, Campbell RK, West KP. Association between stunting and early childhood development among children aged 36–59 months in South Asia. *Matern Child Nutr* 2018;14(Suppl 4):e12684.
- [29] McCoy DC, Peet ED, Ezzati M, Danaei G, Black MM, Sudfeld CR, et al. Early childhood developmental status in low- and middle-income countries: national, regional, and global prevalence estimates using predictive modeling. *PLoS Med* 2016;13:e1002034.
- [30] Tran TD, Holton S, Nguyen H, Fisher J. Physical growth: is it a good indicator of development in early childhood in low- and middle-income countries? *BMC Pediatr* 2019;19:276.
- [31] Rubio-Codina M, Attanasio O, Grantham-McGregor S. Mediating pathways in the socio-economic gradient of child development: evidence from children 6–42 months in Bogota. *Int J Behav Dev* 2016;40:483–91.
- [32] Walker SP, Chang SM, Powell CA, Grantham-McGregor SM. Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. *Lancet* 2005;366:1804–7.
- [33] Yousafzai AK, Obradović J, Rasheed MA, Rizvi A, Portilla XA, Tirado-Strayer N, et al. Effects of responsive stimulation and nutrition interventions on children's development and growth at age 4 years in a disadvantaged population in Pakistan: a longitudinal follow-up of a cluster-randomised factorial effectiveness trial. *Lancet Glob Health* 2016;4:e548–58.
- [34] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.