ORIGINAL ARTICLE

Validation of a touchscreen assessment tool to screen for cognitive delay at 24 months

Tríona Casey[1,2] | **Aoife Jones Thachuthara**[1,2] | **Leanna Fogarty**[1,2] | **Vicki Livingstone**[1,2] | **Michelle De Haan**[3] | **Neil Marlow**[4] | **Mairead E. Kiely**[1,5] | **Deirdre M. Murray**[1,2]

[1]The Irish Centre for Maternal and Child Health Research (INFANT), University College Cork, Cork, Ireland

[2]Department of Paediatrics and Child Health, University College Cork, Cork, Ireland

[3]Department of Developmental Neurosciences, Institute of Child Health, London, UK

[4]UCL EGA Institute for Women's Health, University College London, London, UK

[5]School of Food and Nutritional Sciences, University College Cork, Cork, Ireland

Correspondence
Deirdre M Murray, Department of Paediatrics and Child Health, University College Cork, Cork, Ireland.
Email: d.murray@ucc.ie

Funding information
Science Foundation Ireland, Grant/Award Number: INFANT/B3067

Abstract

Aim: To validate a touchscreen assessment as a screening tool for mild cognitive delay in typically developing children aged 24 months.

Method: Secondary analysis of data was completed from an observational birth cohort study (The Cork Nutrition & Microbiome Maternal–Infant Cohort Study [COMBINE]), with children born between 2015 and 2017. Outcome data were collected at 24 months of age, at the INFANT Research Centre, Ireland. Outcomes were the Bayley Scales of Infant and Toddler Development, Third Edition cognitive composite score and a language-free, touchscreen-based cognitive measure (Babyscreen).

Results: A total of 101 children (47 females, 54 males) aged 24 months (mean = 24.25, SD = 0.22) were included. Cognitive composite scores correlated with the total number of Babyscreen tasks completed, with moderate concurrent validity ($r = 0.358$, $p < 0.001$). Children with cognitive composite scores lower than 90 (1 SD below the mean, defined as mild cognitive delay) had lower mean Babyscreen scores than those with cognitive scores equal to or greater than 90 (8.50 [SD = 4.89] vs 12.61 [SD = 3.68], $p = 0.001$). The area under the receiver operating characteristic curve for the prediction of a cognitive composite score less than 90 was 0.75 (95% confidence interval = 0.59–0.91; $p = 0.006$). Babyscreen scores less than 7 were equivalent to less than the 10th centile and identified children with mild cognitive delay with 50% sensitivity and 93% specificity.

Interpretation: Our 15-minute, language-free touchscreen tool could reasonably identify mild cognitive delay among typically developing children.

Early identification of cognitive delay is important to increase early intervention for at-risk children. Such identification is largely reliant on parent-report questionnaires, such as the Parent Report of Children's Abilities, Revised[1] or the Ages and Stages Questionnaire.[2] Depending on the screening results, children may have the opportunity to complete full developmental assessments conducted by trained assessors. These include the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III),[3] which has strong language, administrator, and cultural biases.[4] Recent advances in touchscreen technology have the potential to be harnessed for neurodevelopmental assessment that is more child-driven and less administrator-dependent than parent-report questionnaires or full developmental batteries.[5,6] This is increasingly feasible as a testing tool, given growing empirical evidence that children acquire the fine motor skills required for touchscreen use before preschool years, across languages and cultures, including Ireland, the UK, Spain, and Malawi.[6–8] The use of a lightweight portable touchscreen assessment may offer more accessible cognitive testing for children with mobility, motor, or sensory issues, who may currently struggle to engage in physically or sensory-demanding tasks typically involved in developmental batteries.[3,9]

Abbreviation: Bayley-III Bayley Scales of Infant and Toddler Development, Third Edition

Our previous research indicated the feasibility of the Babyscreen touchscreen tool (versions 1.0 and 1.5; Hello Labs, Guildford, UK) with children aged between 18 and 36 months.[4,10] Importantly, these early data included children identified as at risk due to complications at birth.[10] However, the predictive ability of developmental delay is higher in high-risk groups. Our current research focuses on using this new assessment tool as a screening tool for cognitive delay in a typically developing, low-risk cohort. The Babyscreen tool is a valuable opportunity for clinicians to screen children for cognitive delay using fewer resources and in a more time-efficient way. In a clinical setting, the use of Babyscreen as a screening assessment would allow more children to be screened for cognitive delay and inform the decision to conduct a full developmental assessment or not.

To prepare Babyscreen for clinical use as a screening tool, further data are needed to validate the application at the age range commonly used for high-risk developmental follow-up in clinical practice, that is, 24 months of age.[11,12] In the UK, National Institute for Health and Care Excellence guidance recommends a face-to-face assessment at 2 years for children born preterm[13] but also using the Parent Report of Children's Abilities, Revised[1] or a suitable alternative parent-report tool, which tend to be highly language dependent. This reinforces the importance of developing a language-free cognitive screening tool at this age.

Our aims were to validate the Babyscreen application as a screening tool for mild cognitive delay. We wished to evaluate its concurrent validity with the Bayley-III cognitive composite score in a low-risk cohort at 24 months of age. We investigated the optimum cut-off for Babyscreen as a cognitive screening tool for children deemed at risk of mild cognitive delay, as indexed by children's Bayley-III cognitive scores, and present normative reference ranges for performance using the application. Finally, we explored the relationship between the Babyscreen score and the neurodevelopmental constructs assessed by the Bayley-III, that is, the cognitive, language, and motor domains.

METHOD

Participants

This study represents a cross-sectional analysis of a sample of 24-month-old children from a birth cohort study: the Cork Nutrition and Microbiome Maternal–Infant Cohort Study (COMBINE). This was a longitudinal, prospective birth cohort study based in Cork, Ireland, running from early pregnancy to 24 months of age, as described in previous research.[14] Inclusion criteria referred to females who were part of the IMPROvED pregnancy cohort study, which included low-risk, nulliparous females with a singleton pregnancy, who attended antenatal care at Cork University Maternity Hospital. Exclusion criteria referred to infants who were admitted to the neonatal intensive care unit for more than 2 weeks, or infants with severe metabolic or congenital anomalies requiring ongoing

What this paper adds

- Babyscreen scores were moderately correlated with Bayley Scales of Infant and Toddler Development, Third Edition cognitive composite scores in typically developing, 24-month-old children.
- A Babyscreen score lower than 7 is the optimal cut-off for identifying mild cognitive delay.
- Babyscreen has the potential to be developed into a 15-minute language-free cognitive screening tool in young children.

specialist care during the neonatal period. The study complied with good clinical practice and had ethical approval from the Cork Research Ethics Committee (ECM4[hh]06/01/15 and ECM3[mmm]25/07/19), with the approved protocol followed. Written informed consent was obtained from all parents. This resulted in an initial sample of 456 mother–infant dyads who participated in the COMBINE study.

The initial sample for this study's secondary data analyses consisted of 134 typically developing children attending for COMBINE neurodevelopmental assessment at 24 months of age. Children were recruited at Cork University Maternity Hospital and were born between 6th January 2016 and 30th November 2017. Mean gestational age was 40.31 weeks (range = 35.29–42.14 weeks). Inclusion criteria for the current study were children aged 24 months who completed both Babyscreen and Bayley-III assessments at the final 24-month appointment (n = 134). Children with disabilities (e.g. motor, visual, hearing, neurodevelopmental, neurological) were not to be excluded from the study; however, it is worth noting that none of the children in the final sample had an identified disability, based on parent report before the 24-month COMBINE appointment. No parents refused the touchscreen assessment, but we excluded children who did not engage independently with the touchscreen (e.g. no interest in assessment, evidence of parent prompting; n = 9), children who did not have a Bayley-III cognitive composite score (n = 3), and children who came from a non-English-speaking household (n = 21). The latter criterion was applied to exclude language effects on the validity of the Bayley-III cognitive score. This left a final sample of 101 children for our non-randomized case cohort study. For details of participant flow, please refer to Figure S1. The demographic characteristics of the Babyscreen study sample were largely reflective of the initial COMBINE cohort (Table S1).

Assessment

Neurodevelopmental follow-up was conducted at the INFANT Research Centre, Cork University Hospital, when the child was aged from 24 months 0 days to 24 months 31 days. Data were collected between 11th January 2018 and 3rd December

2019. Evaluation included a Bayley-III and Babyscreen touchscreen assessment, both described in the following sections. Assessments were performed by study site professionals proficient in the administration of the respective tools.

Instruments

The Babyscreen software application v1.89 (Hello Games, Guildford, UK) is an assessment tool designed to tap into toddlers' fundamental cognitive abilities. Previous work showed that typically developing and at-risk children can interact meaningfully with Babyscreen.[4,10] The Babyscreen application was used on a fully charged Apple iPad 2.0 ($241.2 \times 185.7 \times 8.8$ mm), with protective rubber covering to prevent damage. The iPad screen was cleaned with child-friendly disinfectant wipes pre- and post-assessment. Given that the results of the Babyscreen assessment were transferred from the iPad to the lab computer directly, no paper was needed, thereby removing the cost of printing materials. For this study, one performance variable of Babyscreen was focused on the number of items completed successfully (range 0–18), based on the total number of tasks completed independently by the child, without needing verbal instruction or a visual demonstration from the administrator, within a time limit of 30 seconds per task. No receptive or expressive language is required by the child to complete the Babyscreen assessment. Figure S2 displays the schematic of the 18 tasks used in the Babyscreen assessment, categorized under the respective cognitive constructs.

The Bayley-III was developed and validated in a US population and is designed to quantify the developmental functioning of infants and children from 1 to 42 months of age. It consists of cognitive, language (receptive and expressive), and motor (fine and gross) scales. Each scale is converted into a composite score that is standardized to a mean of 100 and an SD of 15. Scores of 85 to 115 are within the average range and scores below 85 are indicative of cognitive delay. However, there are concerns over the standardization of the Bayley-III and the use of US norms in different countries, given widespread cross-cultural evidence for the overestimation of cognitive function using Bayley-III norms.[15–21] The present study followed the recommendation of using a cut-off for mild cognitive delay of less than 1 SD below the mean of a geographically relevant control group, to improve the detection of developmental delay.[15,22] Previous researchers reported this cut-off to fall around 90, based on a mean cognitive composite score of approximately 105 and an SD of 15.[10,22]

Statistical analysis

All data analysis was carried out using SPSS v26 (IBM Corp., Armonk, NY, USA). Descriptive data analyses are presented as the mean (SD) as appropriate. An alpha value of 0.05 was used to determine statistical significance. Because Babyscreen and Bayley-III scores were normally distributed, the relationship between Babyscreen total and Bayley-III cognitive composite

scores was assessed using a Pearson's correlation coefficient. A Student's t-test was used to investigate the difference in Babyscreen and Bayley-III cognitive composite scores for those scoring above and below the cognitive cut-off. The ability of the Babyscreen score to predict a child's cognitive composite score below the cut-off was evaluated using receiver operating characteristic curves. The sensitivity, specificity, and predictive values corresponding to the optimal Babyscreen cut-off obtained from the receiver operating characteristic curve analysis were calculated. The intercorrelations between the Bayley-III subscales and the two target variables (Babyscreen score and Bayley-III cognitive composite score) were calculated using the Pearson's correlation coefficient; t-tests and analysis of variance (ANOVA) were used to test for the effect of confounding variables on the Babyscreen score. Observed centiles for performance on the Babyscreen were produced to indicate normative reference ranges (using scores of all children aged 24 months who engaged with the assessment tool, including those from non-English-speaking households). Power analyses were completed with MedCalc v20.116 (MedCalc Software, Ostend, Belgium),[23] based on a two-tailed alpha of 0.05 and beta of 0.20. Our sample size of 101 fulfilled all criteria for sample size and ratio of positive to negative cases for cognitive delay. Missing data for the Babyscreen score or Bayley-III cognitive composite scores were managed via pairwise exclusion.

RESULTS

Sample

One hundred and thirty-four children were assessed using both Bayley-III and Babyscreen assessments at 24 months of age. From these, 33 cases were excluded, as described previously. This left a final study sample of 101 children aged 24 months. The 101 children (47 females, 54 males) were born with a mean gestational age of 40.27 weeks (range 35.29–42.14 weeks) and a mean birth weight of 3632 g (SD = 459). Children were followed up at 24 months of age (mean = 24.25, SD = 0.22). Ninety-nine mothers self-identified their nationality as Irish (two British), with 99 self-identifying as White Irish (two White non-Irish). The three children with missing data for the cognitive composite score were excluded. Some children had no language composite (6) or motor composite (9) scores due to tiredness on the day of testing. However, these were deemed to be secondary variables; thus, analyses proceeded as planned.

Confounding variables

Several variables were tested for potential effects on the Babyscreen score, including the sex of the child, touchscreen use, and several demographic and socioeconomic proxy indicators (Table 1). There were no significant differences in Babyscreen scores for any of the potential confounding variables. Therefore, we did not control for these variables in further analyses.

Concurrent validity of Babyscreen and Bayley-III cognitive composite score

There was a significant moderate positive correlation between the Bayley-III cognitive composite score and the number of Babyscreen tasks completed (n = 101, r = 0.358, $p < 0.001$) as categorized by Cohen's criteria.[24]

Babyscreen scores of children with and without cognitive delay

Our sample had a mean (SD) Babyscreen total score of 12 (SD = 4) items and a mean (SD) Bayley-III cognitive composite score of 104 (SD = 15.7). A score lower than 90 was used to represent mild cognitive delay, corresponding to approximately 1 SD below our sample mean. Eighty-nine children had a cognitive composite score equal to or greater than 90 and 12 scored less than 90. Children with a cognitive score lower than 90 completed an average of 8.5 (SD = 4.9) Babyscreen tasks compared with those with a cognitive score equal to or greater than 90 who completed an average of 12.6 (SD = 3.7) Babyscreen tasks (t[99] = −3.484, p = 0.001). The magnitude of the difference in the means was moderate (mean difference = 4.11; 95% confidence interval [CI] = −6.445 to −1.768; η^2 = 0.109). The distribution of Babyscreen scores for children with and without mild cognitive delay is shown in Figure 1.

Predictive ability of Babyscreen tool in identifying the risk of cognitive delay

Receiver operating characteristic curve analyses indicated that the Babyscreen score could predict the cognitive delay indicated by a Bayley-III cognitive composite score lower than 90 (p = 0.006, area under the curve = 0.746, 95% CI = 0.59–0.91). The optimal Babyscreen cut-off score for maximizing sensitivity and specificity was 6.5, with a Youden index of 0.433. This cut-off yielded a sensitivity of 50% and a specificity of 93% in predicting a cognitive score lower than 90. The positive predictive value of the Babyscreen cut-off was 50% (95% CI = 27.73–72.27), while the negative predictive value was 93.3% (95% CI = 88.68–96.07). Of the 89 children with a Babyscreen score equal to or greater than 7, six (6.7%) had a Bayley-III cognitive composite score lower than 90 compared with the 12 children with a Babyscreen total lower than 7, of whom six (50%) had a cognitive composite score lower than 90.

T A B L E 1 Comparison[a] of mean (SD)[b] Babyscreen scores based on demographics and socioeconomic indicators.

Variable	n	Mean (SD)	Difference in means (95% CI)[c]	p
Child sex				
Male	54	11.98 (4.49)	0.295 (−1.31 to 1.90)	0.716
Female	47	12.28 (3.50)		
Admittance to neonatal unit				
Admitted	13	12.46 (4.24)	0.393 (−2.00 to 2.79)	0.745
Not admitted	88	12.07 (4.03)		
Previous touchscreen experience				
Never	15	13.47 (4.41)		0.425
Occasionally	44	12.27 (4.08)		
2–3 times per week	17	11.35 (3.72)		
Daily	25	11.56 (4.00)		
Maternal education				
Secondary education	28	11.21 (4.38)		0.380
Undergraduate education	39	12.41 (4.15)		
Postgraduate education	34	12.53 (3.60)		
Combined household income				
<€21K–42K	27	11.96 (4.00)		0.934
€43K–84K	46	12.28 (3.98)		
>€85K	28	12.00 (4.31)		
Maternal employment status				
Yes, employed	74	12.26 (3.93)		0.740
Yes, on maternity leave	18	11.44 (4.69)		
Not employed	9	12.33 (3.87)		

[a]Based on the t-test or analysis of variance results.

[b]Mean number of tasks completed independently by the child within 30 seconds, without assessor instruction or visual demonstration (score range 0–18).

[c]Mean difference and 95% confidence interval (CI) reported for t-tests.
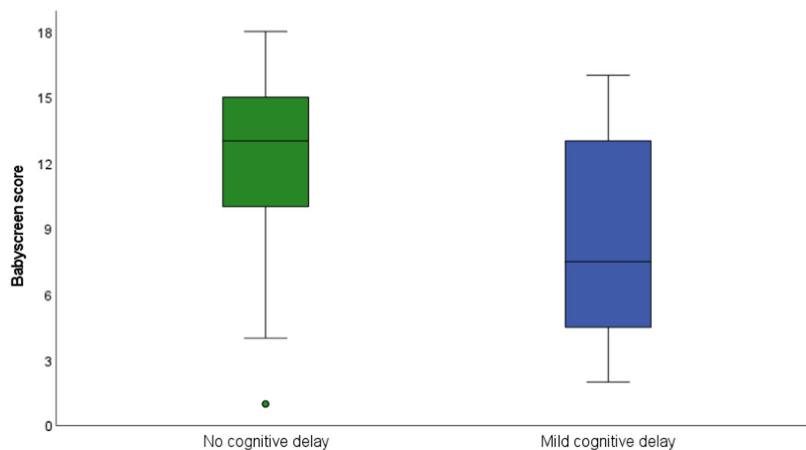
FIGURE 1 Box plot of Babyscreen scores for children with normal Bayley Scales of Infant and Toddler Development, Third Edition cognitive composite scores equal to or greater than 90 (*n* = 89) (green) and children with a cognitive score lower than 90 (*n* = 12) (blue). The Babyscreen score reflects the number of tasks completed independently by the child within 30 seconds, without assessor instruction or visual demonstration (score range 0–18).

TABLE 2 Pearson's correlations (*p*) between Bayley-III subtests and the two performance variables, the Babyscreen score and Bayley-III cognitive score.

	Babyscreen score[a] (*n* = 101)	Bayley-III cognitive score[b] (*n* = 101)
Bayley-III cognitive composite score	0.358 (*p* < 0.001)	N/A
Bayley-III language composite score	0.144 (*p* = 0.163)	0.722 (*p* < 0.001)
Receptive language scaled score	0.149 (*p* = 0.148)	0.706 (*p* < 0.001)
Expressive language scaled score	0.117 (*p* = 0.256)	0.638 (*p* < 0.001)
Bayley-III motor composite score	0.249 (*p* = 0.016)	0.568 (*p* < 0.001)
Fine motor scaled score	0.253 (*p* = 0.014)	0.655 (*p* < 0.001)
Gross motor scaled score	0.192 (*p* = 0.066)	0.299 (*p* = 0.004)

[a]The score reflects the number of Babyscreen tasks completed independently by the child within 30 seconds, without assessor instruction or visual demonstration (score range 0–18).
[b]The score reflects the Bayley-III cognitive composite score (score range 55–145).
Abbreviation: Bayley-III, Bayley Scales of Infant and Toddler Development, Third Edition.

Relationships of the Babyscreen and Bayley-III cognitive scores with the Bayley-III developmental subscales

We calculated the correlations between the Babyscreen score and remaining composite and scaled scores (Table 2). The Babyscreen score did not correlate with the language composite score (*n* = 95, *r* = 0.144, *p* = 0.163) and had a weak correlation with the motor composite score (*n* = 92, *r* = 0.249, *p* = 0.016). The latter correlation was driven by the fine motor subscale, which showed a weak positive correlation with the

Babyscreen score (*n* = 94, *r* = 0.253, *p* = 0.014). We also examined the intercorrelations between Bayley-III subscales. The cognitive composite score correlated strongly with both language (*n* = 95, *r* = 0.722, *p* < 0.001) and motor (*n* = 92, *r* = 0.568, *p* < 0.001) composites. The relationship between cognitive and motor composite scores was driven by the fine motor subtest, which recorded a strong correlation with the cognitive composite score (*n* = 94, *r* = 0.655, *p* < 0.001).

Reference values

The overall histogram and calculated centiles for performance on the Babyscreen assessment for all children tested at 24 months are shown in Figure 2. Note that this normative data set was extended from 101 to 125 children (i.e. including 21 children from non-English-speaking households and three children without a Bayley-III cognitive composite score but excluding nine children who did not interact independently with the application). The estimated cut-off for the detection of cognitive delay (<7) is equivalent to the 10th centile for the Babyscreen performance.

DISCUSSION

In a typically developing cohort aged 24 months, the Babyscreen total score showed moderate concurrent validity with the Bayley-III cognitive composite score and provided moderate predictive value for cognitive delay. Babyscreen performance was independent of a range of socioeconomic measures and previous touchscreen use.

Our results build on previous evidence that demonstrated similarly reasonable concurrent validity between Babyscreen and the Bayley-III cognitive subscale in a high-risk cohort aged between 18 and 24 months.[10] Our previous study suggested that Babyscreen had potential as a cognitive screening tool for children deemed at risk of cognitive delay. We
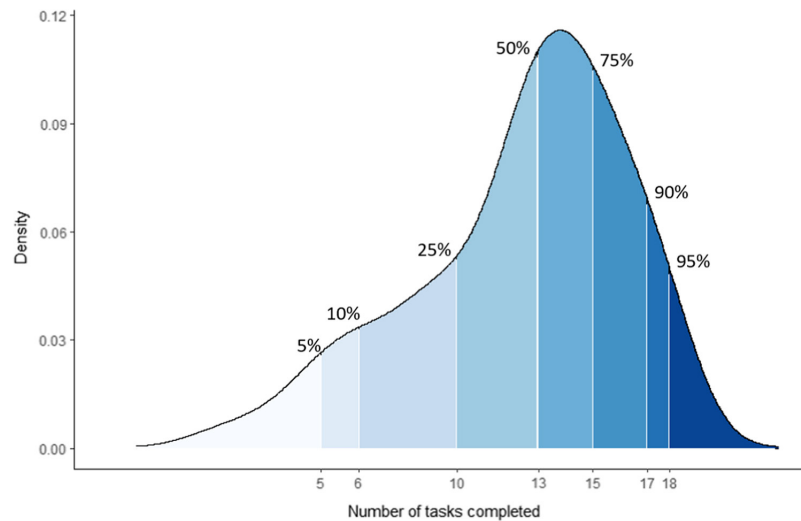
FIGURE 2　Histogram and centile ranges of Babyscreen scores for all children aged 24 months who engaged independently with the assessment tool (*n* = 125). Centile ranges are indicated on the graph (5% = 5th centile). The Babyscreen score reflects the number of tasks completed independently by the child within 30 seconds, without assessor instruction or visual demonstration (score range 0–18).

extended this research by validating Babyscreen as a screening tool in a low-risk cohort aged 24 months. Using the screening cut-off identified with these data, children who achieved a Babyscreen score lower than 7 were more likely to attain a low cognitive composite score on the Bayley-III and thus would be recommended for a full developmental assessment. This research holds important translational implications for developmental follow-up of children. Babyscreen is a quick and practical option for clinical practitioners to screen large numbers of children aged 24 months for cognitive delay. This could reduce clinical caseloads by excluding typically developing children and highlighting those at risk, enabling quicker access to full developmental assessment and earlier intervention for the latter.

The study also explored the neurodevelopmental constructs being tapped into by the Babyscreen and Bayley-III assessments (cognitive, language, and motor domains). There was no association between the Babyscreen score and Bayley-III language composite score. This contrasts with the strong association between Bayley-III cognitive and language composite scores. The strong overlap between language and cognitive scales in the Bayley-III assessment puts children from non-English-speaking households at a disadvantage. The Babyscreen score had a weak association with the Bayley-III motor composite score, which was driven by the fine motor subscale. This is to be expected, given the link between cognitive and fine motor coordination as seen in the strong relationship between Bayley-III cognitive and fine motor scores in the current sample and as previously reported by the Bayley-III developers.[25] Importantly, Babyscreen scores were most strongly correlated with Bayley-III cognitive composite scores. This indicates that the Babyscreen assessment targets the cognitive neurodevelopmental domain independently from the language domain, thus underlying its value as a cognitive screening tool. We compared it to the Bayley-III in this study because it was conducted before

the introduction of the Bayley Scales of Infant and Toddler Development, Fourth Edition in 2019. However, we expect that similar findings will apply because of the strong correlation reported between the Bayley-III and Bayley Scales of Infant and Toddler Development, Fourth Edition.[26]

Feasibility of the Babyscreen assessment

The feasibility of using Babyscreen as a cognitive screening tool is highly relevant given the benefits of using touchscreen assessment tools with children. While the Bayley-III is the most popular developmental assessment, it is not without faults, including a lengthy administration time and heavy reliance on the child's language ability. A lightweight, transportable touchscreen device makes it more feasible for assessments by eliminating large testing kits and potentially improving test accessibility for children with mobility issues.[6] Furthermore, by the age of 24 months, children have developed an array of motor skills required to use touchscreen technology.[7,8] The use of touchscreen assessments would also promote standardized testing due to reduced administrator interaction, thus decreasing tester bias and measurement error.[6] Our results support the call for greater use of non-verbal cognitive assessments to gain a more accurate reflection of young children's cognitive abilities, especially for children with speech delay or difficulties and for whom English is not their first language.[4,5] The use of this language-free, touchscreen-based assessment offers a feasible option for cognitive screening by experienced psychologists and medical practitioners, particularly in our increasingly multicultural societies.

Given the ubiquity of touchscreen devices today, it is important to make the distinction between using touchscreens as an assessment tool and for other uses. Although

our analyses did not find any difference in the scores of children with different reported levels of touchscreen use, most children in our geographical area are exposed to touchscreen devices from a young age.[8] Conflicting evidence exists regarding the long-term developmental effect of children's exposure to touchscreen devices,[27,28] which is probably related to the quality of content, interactivity and engagement, and quantity.[29,30] It is essential to use Babyscreen solely as an assessment tool, not for entertainment or other purposes.

Limitations

Our analyses are restricted to a 24-month sample to validate this cognitive screening tool at a commonly used time point of developmental follow-up. Further testing is required to establish normative data and cut-off scores for other age ranges. Our sample included predominately White Irish participants, with relatively high education and income compared to other countries. Our newly reported normative data provide a starting point to validate Babyscreen with multinational, multi-ethnic, and multilingual and other-lingual participants in Ireland and further afield.

Our sample was derived from a typically developing cohort of children, with the final sample having no children with identified physical, visual, or hearing difficulties, or with any known neurodevelopmental or neurological conditions. However, our study relied on parent report before the 24-month assessment to indicate if their child had a disability, instead of using criterion standard clinical or medical examinations to assess for the presence or absence of disabilities. It was not possible to assess the accessibility of Babyscreen for children with specific disabilities (e.g. visuo-motor, fine motor, somatosensory processing, or social communication difficulties). Further research should explore the feasibility of Babyscreen as a cognitive screening tool for children with different physical, sensory, or neurological disabilities.

A relatively small number of children were identified as having mild cognitive delay within this typically developing sample. Therefore, while the specificity of the Babyscreen tool is relatively strong, it was not possible to get better tool sensitivity. For example, children with Bayley-III cognitive scores lower than 90 who completed more than seven tasks on the Babyscreen assessment may have cognitive skills within the normal range but may have underperformed on the Bayley-III due to attentional or behavioural difficulties on the day. Furthermore, while we confirmed the ability of Babyscreen as a screening tool for cognitive delay with a binary outcome recommendation for developmental follow-up, we do not yet have the data to recommend its use as an independent cognitive assessment.

Finally, we acknowledge that our results have limited generalizability. Although our sample is relatively similar demographically to the parent birth cohort study, the generalizability of findings is hampered because of the non-randomized design of our study.

Future research

Future research should assess larger numbers of children from typically developing and at-risk cohorts. A larger sample size may improve the overall strength of the moderate correlation found between Babyscreen and Bayley-III, while also allowing for more accurate comparison of results between participant subgroups (e.g. previous touchscreen use, ethnicity, disability). Future testing and analyses would thus guide future development as a complete cognitive assessment encompassing relevant neuropsychological domains (Figure S1). Future research could compare the accuracy of Babyscreen as a cognitive screening tool with parent-report questionnaires,[1,2] given that Babyscreen may be less vulnerable to parental, language, and cultural bias. It is also important to assess Babyscreen's predictive validity for identifying children who have a definitive disability or neurodevelopmental condition at school age. Given the current cross-sectional analyses, evaluation of Babyscreen's ability to predict children's future cognitive scores should be investigated through further longitudinal studies.

To assess the generalizability of Babyscreen, future studies should consider the acceptability and validity of this touchscreen assessment if implemented in a general population (e.g. alongside public health developmental check-ups). Additional testing of Babyscreen across different low- and high-income countries is required to gain more cross-cultural relevance, given that rates of touchscreen use are likely to differ.[4,6] There is already promising research that indicates the reliability and validity of using touchscreen assessments to assess cognitive skills in school-aged children in Malawi and in the UK.[6] Future research will need to address the potential effect of touchscreen exposure and cross-cultural variances on the viability of Babyscreen with similarly aged cohorts in different cultural settings.

Conclusions

We have shown that the Babyscreen touchscreen assessment has potential to be developed as a screening tool for early mild cognitive delay, defined as 1 SD below the sample mean on the Bayley-III cognitive composite score. This language-independent tool, previously tested with high-risk children, was suitable for use with typically developing 24-month-old children. We have reported reference values for performance in a typically developing cohort.

DATA AVAILABILITY STATEMENT
Data are available on reasonable request.

ORCID
Tríona Casey https://orcid.org/0000-0002-7789-9072
Neil Marlow https://orcid.org/0000-0001-5890-2953

REFERENCES
1. Johnson S, Bountziouka V, Brocklehurst P, Linsell L, Marlow N, Wolke D, et al. Standardisation of the Parent Report of Children's Abilities-Revised (PARCA-R): a norm-referenced assessment of cognitive and language development at age 2 years. Lancet Child Adolesc Health. 2019;3(10):705–12.
2. Singh A, Yeh CJ, Boone Blanchard S. Ages and Stages Questionnaire: A global screening scale. Boletín Médico Del Hospital Infantil de México (English Edition). 2017;74(1):5–12.
3. Bayley N. Bayley scales of infant and toddler development, third edition: administration manual. San Antonio, USA: Harcourt Assessment; 2006.
4. Twomey DM, Wrigley C, Ahearne C, Murphy R, De Haan M, Marlow N, et al. Feasibility of using touch screen technology for early cognitive assessment in children. Arch Dis Child. 2018;103(9):853–8.
5. Dan B, Pelc K. Application of touch screen technology for early assessment of executive function. Arch Dis Child. 2019;104(2):108–9.
6. Pitchford NJ, Outhwaite LA. Can touch screen tablets be used to assess cognitive and motor skills in early years primary school children? A cross-cultural study. Front Psychol. 2016;7:1666.
7. Nacher V, Jaen J, Navarro E, Catala A, González P. Multi-touch gestures for pre-kindergarten children. Int J Hum Comput Stud. 2015;73:37–51.
8. Ahearne C, Dilworth S, Rollings R, Livingstone V, Murray D. Touch-screen technology usage in toddlers. Arch Dis Child. 2016;101(2):181–3.
9. Stroud L, Green E, Bloomfield S, McAlinden P. Griffiths III - Revision process, standardisation and psychometric properties and a link to neurorehabilitation. European Journal of Paediatric Neurology. 2017;21:e145.
10. Twomey DM, Ahearne C, Hennessy E, Wrigley C, De Haan M, Marlow N, et al. Concurrent validity of a touchscreen application to detect early cognitive delay. Arch Dis Child. 2021;106(5):504–6.
11. Johnson S, Marlow N. Developmental screen or developmental testing? Early Hum Dev. 2006;82(3):173–83.
12. Johnson S, Marlow N. Early and long-term outcome of infants born extremely preterm. Arch Dis Child. 2017;102(1):97–102.
13. National Institute for Health and Care Excellence. Developmental follow-up of children and young people born preterm. 2017. Available from: https://www.nice.org.uk/guidance/ng72/resources/developmental-followup-of-children-and-young-people-born-preterm-pdf-1837630868677 (date accessed 5th April 2020).
14. Hemmingway A, Fisher D, Berkery T, Dempsey E, Murray DM, Kiely ME. A detailed exploration of early infant milk feeding in a prospective birth cohort study in Ireland: combination feeding of breast milk and infant formula and early breast-feeding cessation. British Journal of Nutrition. 2020;124(4):440–9.
15. Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? Pediatr Res. 2014;75(5):670–4.
16. Fuiko R, Oberleitner-Leeb C, Klebermass-Schrehof K, Berger A, Brandstetter S, Giordano V. The impact of norms on the outcome of children born very-preterm when using the Bayley-III: differences between US and German norms. Neonatology. 2019;116(1):29–36.
17. Li F, Wu SS, Berseth CL, Harris CL, Richards JD, Wampler JL, et al. Improved neurodevelopmental outcomes associated with bovine milk fat globule membrane and lactoferrin in infant formula: a randomized, controlled trial. J Pediatr. 2019;215:24–31.
18. Chinta S, Walker K, Halliday R, Loughran-Fowlds A, Badawi N. A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). Arch Dis Child. 2014;99(7):621–4.
19. Steenis LJP, Verhoeven M, Hessen DJ, van Baar AL. Performance of Dutch children on the Bayley III: a comparison study of US and Dutch Norms. PLoS One. 2015;10(8):1–13.
20. Gasparini C, Caravale B, Rea M, Coletti MF, Tonchei V, Bucci S, et al. Neurodevelopmental outcome of Italian preterm children at 1year of corrected age by Bayley-III scales: an assessment using local norms. Early Hum Dev. 2017;113:1–6.
21. Çelik P, Ayranci Sucakli I, Yakut HI. Which Bayley-III cut-off values should be used in different developmental levels? Turk J Med Sci. 2020;50(4):764–70.
22. Aylward GP. Continuing issues with the Bayley-III: where to go from here. J Dev Behav Pediatr. 2013;34(9):697–701.
23. MedCalc. MedCalc Statistical Software. 2016.
24. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates; 1988.
25. Bayley N. Bayley scales of infant and toddler development, third edition: technical manual. San Antonio, USA: Harcourt Assessment; 2006.
26. Balasundaram P, Avulakunta ID. Bayley Scales Of Infant and Toddler Development. [Updated 2021 Nov 24]. 2022 Jan-. In: StatPearls [Internet] [Internet]. StatPearls Publishing: Treasure Island (FL). Available from: https://www.ncbi.nlm.nih.gov/books/NBK567715/.
27. World Health Organization. Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age. 2019. Available from: https://www.who.int/publications-detail/guidelines-on-physical-activity-sedentary-behaviour-and-sleep-for-children-under-5-years-of-age (date accessed 5th April 2020).
28. Eirich R, McArthur BA, Anhorn C, McGuinness C, Christakis DA, Madigan S. Association of screen time with internalizing and externalizing behavior problems in children 12 years or younger: a systematic review and meta-analysis. JAMA Psychiat. 2022;79(5):393–405.
29. Christakis DA, Garrison MM, Herrenkohl T, Haggerty K, Rivara FP, Zhou C, et al. Modifying media content for preschool children: a randomized controlled trial. Pediatrics. 2013;131(3):431–8.

30. Madigan S, McArthur BA, Anhorn C, Eirich R, Christakis DA. Associations between screen use and child language skills: a systematic review and meta-analysis. JAMA Pediatr. 2020;174(7):665–75.

SUPPORTING INFORMATION

The following additional material may be found online:

Table S1 Comparison of demographic characteristics between initial sample of mother–infant dyads in the COMBINE birth cohort study and the Babyscreen study sample at 24 months

Figure S1 Participant flow for the 24-month Babyscreen study.

Figure S2 Schematic of 18 tasks on the Babyscreen touchscreen assessment.

How to cite this article: Casey T, Thachuthara AJ, Fogarty L, Livingstone V, De Haan M, Marlow N, et al. Validation of a touchscreen assessment tool to screen for cognitive delay at 24 months. Dev Med Child Neurol. 2023;00:1–9. https://doi.org/10.1111/dmcn.15555