Concurrent validity of a touchscreen application to detect early cognitive delay

Deirdre Marie Twomey,[1,2] Caroline Ahearne,[1,2] Emma Hennessy,[2,3] Conal Wrigley,[2,3] Michelle De Haan,[4] Neil Marlow [ORCID],[5] Deirdre M Murray [ORCID] [1,2]

[1]Department of Paediatrics and Child Health, University College Cork National University of Ireland, Cork, Ireland
[2]The Irish Centre for Maternal and Child Health Research, University College Cork, Cork, Ireland
[3]Department of Applied Psychology, University College Cork National University of Ireland, Cork, Ireland
[4]Dept of Developmental Neurosciences, Insitute of Child Health, University College London, London, UK
[5]Institute for Women's Health, University College London, London, UK

Correspondence to
Prof Deirdre M Murray, Dept of Paediatrics and Child Health, University College Cork, Cork T12 K8AF, Ireland; d.murray@ucc.ie

ABSTRACT

Objective To explore the ability of an interactive screening tool to identify cognitive delay in children aged 18 to 24 months.

Design Children were assessed using the Bayley Scale of Infant and Toddler Development—third edition (BSID-III) and a touchscreen measure of problem-solving (Babyscreen V.1.5). We examined the internal consistency and concurrent validity between the two measures. A BSID-III cognitive composite score (BSID-IIIcc) ≤1 SD below population mean was used to indicate a low average cognitive ability.

Results 87 children with a mean (SD) age of 20.4 (1.3) months who experienced complications at delivery (n=53) and healthy age-matched controls (n=34) were included in the study. A moderate positive correlation between the BSID-IIIcc and the total number of tasks completed on the Babyscreen suggested reasonable concurrent validity (r=0.414, p<0.001). Children with a BSID-IIIcc ≤90 had lower median (IQR) Babyscreen score (7 (6, 8.5) vs 11 (8.5, 13); p=0.003) and a lower median (IQR) age-adjusted z-score (BST z-score) for number of items completed compared with those >90 (−1.08 (−1.5 to −0.46) vs 0.31 (−0.46 to 0.76); p=0.001). The area under the receiver operating characteristic curve for the prediction of a low normal BSID-IIIcc was 0.787 (CI 0.64 to 0.93). A BST z-score of <−0.44 yielded 82.4% sensitivity and 71.4% specificity in identifying children with cognitive delay.

Conclusions A touchscreen-based application has concurrent validity with the BSID-IIIcc and could be used to screen for cognitive delay at 18–24 months of age.

What is already known?

► Cognitive delay is difficult to detect in pre-verbal children.
► Standard administered developmental assessments are heavily language dependent.
► From 12 months of age, children can interact in a meaningful way with touchscreen applications.

What this study adds?

► A touchscreen application designed to test problem-solving ability has good concurrent validity with the Bayley Scale of Infant and Toddler Development—third edition cognitive composite score.
► Performance on the Babyscreen app has potential to screen for low average cognitive ability at 18–24 months.

Our ability to assess cognitive development in early childhood is limited by the use of proxy measures of cognitive function, such as parental report questionnaires, or developmental assessments, such as the Bayley Scale of Infant and Toddler Development (BSID-III).[1] These assessment tools require a trained administrator, and rely heavily on the child's verbal and motor skills. Intraobserver variability is also significant.[2] The majority of current assessment tools have been validated only in English-speaking populations and are therefore not appropriate for use in children with a non-English mother tongue.

Rapid progress in the field of computerised interactive technology can facilitate pre-verbal children engaging in tasks without verbal instruction.[3] We recently demonstrated that children as young as 24 months can complete a measure of problem-solving presented on a touch screen device without verbal instruction.[4] In this paper, we report the concurrent validity of the Babyscreen V.1.5 and the BSID-III in children aged 18–24 months. We also aimed to assess the ability of the Babyscreen to detect cognitive delay as defined by the BSID-IIIcc and examine if a performance cut-off could be used to aid screening.

METHODS

The study sample comprised children attending for neurodevelopmental assessment between 18 and 24 months at Cork University Maternity Hospital, Cork, Ireland. Neurodevelopmental assessment comprised the BSID-III[1] and the Babyscreen V.1.5. The Babyscreen Software Application V.1.5 (Hello Games, UK) is an 18-item cognitive assessment tool designed to tap into basic cognitive capacities as previously described.[4] Our feasibility study indicated that the total number of items completed without demonstration was most closely related to age and was a useful overall summative measure of problem-solving; denoted as the Babyscreen score (0–18) in the present study.

Statistical analysis

The internal consistency of the Babyscreen was assessed using Cronbach's alpha. Data were normally distributed and correlation was assessed using Pearson's correlation coefficient. We calculated a z-score (BST z-score) for the total number

Table 1 Performance profile of children with a low average cognitive performance (BSID-III cognitive composite score ≤90) at 18–24 months assessed using two testing methods: the BSID-III and a summative measure of performance on the Babyscreen application

	BSID-IIIcc >90 (n=70)	BSID-IIIcc ≤90 (n=17)	P value*
Cognitive composite	105 (100 to 115)	90 (82.5 to 90)	<0.001
Language composite	106 (100 to 117)	97 (90 to 105)	0.003
Motor composite	107 (100 to 112)	97 (94 to 101)	<0.001
Babyscreen score	11 (8.5 to 13)	7 (6 to 8.5)	0.003
BST z-score	0.31 (−0.46 to 0.76)	−1.08 (−1.51 to −0.46)	0.001

Babyscreen score=total number of touchscreen items completed without visual demonstration. BST z-score=age-adjusted Babyscreen z-score grouped by age 18–21 and 22–24 months. Values listed as median (IQR).
*P value=Mann-Whitney U between children with BSID-III >90 and those with BSID-III ≤90.

of Babyscreen items mean (SD) completed in the control group in two age categories: children aged 18–20 months and children aged 21–24 months. The ability of a child's BST z-score to predict a BSID-III ≤90 were assessed using receiver operating characteristic (ROC) curves. The cut-off score of ≤90 was employed as this value corresponds to approximately 1 SD below the cohort mean (M=103; SD=11.9).[5] A Bonferroni correction was applied when several comparisons were performed simultaneously.

RESULTS

Of the 136 children who attended for neurodevelopmental assessment, touchscreen assessment was attempted in 113 children. Technical difficulties (app not recording/data not saved/ data lost) affected 15 assessments so that the data could not be analysed. Five of the 113 children did not engage with the app due to behaviour or were felt to be 'too tired'. Three of these five were of a younger age (19 months) at assessment and were also unable to complete the BSID-III. Therefore, 93 with a median age of 20 months (IQR 19–21) completed both the BSID-III *and* the Babyscreen assessment V.1.5. Of these, six children from non–English-speaking households were excluded from analysis leaving a final study group with complete data of 87 children.

These 87 children (40 females, 47 males) had a median gestational age of 40 weeks (IQR 39–41) and a mean weight of 3585 (SD 504) g. Mean (SD) age at testing was 20.4 (1.3) months.

These children were recruited as controls (n=34), and 53 with signs of perinatal asphyxia at birth, of whom 10 developed HIE. No difference was seen in Babyscreen performance across the categories of previous touchscreen use (none/occasional/2–3 times per week/daily), p=0.773.

The internal consistency of Babyscreen as an 'overall' measure of problem-solving was acceptable as indicated by a Cronbach's alpha value of 0.63. There was a moderate positive association between the cognitive composite score on the BSID-III and the Babyscreen scores for each child (r=0.414, p<0.001). In contrast, the Babyscreen score correlated weakly with the language (r=0.24, p=0.038) and motor (0.28, p=0.019) composite scores of the BSID-III. This indicates that the Babyscreen scores were more closely linked with the cognitive scales of the BSID-III.

Older children (21–24 months, n=41) had significantly higher Babyscreen scores than those aged 18 to 20 months (n=45); mode (range) total number of items completed=12 (10) versus 8 (13), respectively (p=0.011). The age-adjusted BST z-scores also correlated with the BSID-IIIcc (r=0.416, p<0.001). Of the 87 children in the total cohort, the number of children with a BSID-IIIcc ≤90 was 17. There was a significant difference in median (IQR) Babyscreen scores and BST z-scores in those children with a BSID-III cognitive score ≤90 compared with those with a BSID-III cognitive composite score of 91 or greater, indicating a moderate ability to differentiate between those children with no cognitive delay compared with those with low average scores (table 1 and figure 1).

ROC analyses indicated that BST z-score could predict abnormal performance as indexed by a BSID-IIIcc score of ≤90 (p=0.001, AUC=0.787, CI 0.64 to 0.93). The optimal BST-z-score cut-off for maximising both the test sensitivity and specificity was −0.44, with a Youden index of 0.538. This cut-off gave a sensitivity of 82.4% and a specificity of 71% for the prediction of a BSID-IIIcc score of <90. Of the 53 children with had a BST z-score >−0.44, only 3/52 (5.8%) had an abnormal (≤90) BSID-III score. In contrast, in the 34 children with a BST z-score below the cut-off, 14 (41%) had a BSID-III cognitive score ≤90.

DISCUSSION

We have shown that the Babyscreen has reasonable concurrent validity with the cognitive subscale of the BSID-III and that
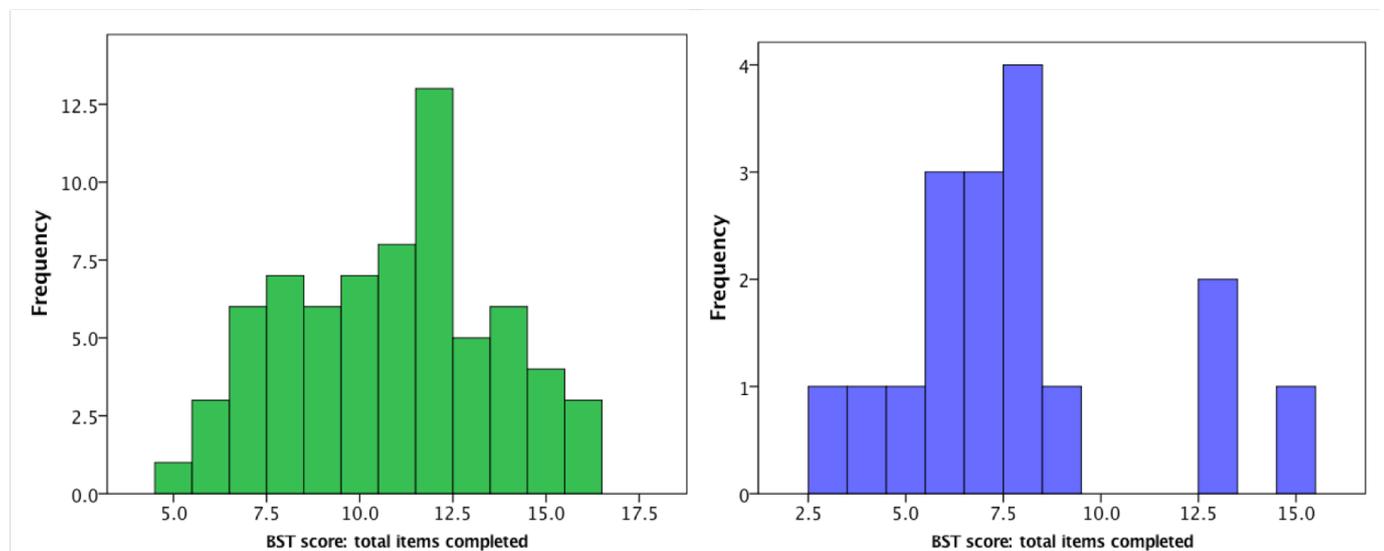
Figure 1 Histograms of Babyscreen scores (total number of items completed without visual demonstration) in those children with normal BSID-III cognitive composite scores (>90, n=71) in light green and in those children (n=17) with a BSID-III cognitive score ≤90 at 18–24 months (teal blue).

performance within an expected range can predict a normal BSID-IIIcc. Performance on the Babyscreen correlated best with the cognitive composite score of the BSID-III and so may focus directly on cognitive aspects of ability.

A screener than can predict normal outcome would be extremely useful, allowing larger populations of children to be monitored quickly and accurately. Although we have taken the BSID-III to be the current most validated measure of developmental delay, its ability to predict later cognitive ability is debatable. We have focused on comparing the two measures, with the knowledge that the BSID-III, while not ideal, is our current most frequently used assessment tool in this age group.

The Babyscreen holds promise for cognitive screening in non–English-speaking and non-verbal children, improving our ability to assess outcome in multicentre studies across multiple countries. We are only beginning to explore the ability of toddlers to engage with complex touchscreen tasks and so these are preliminary data in a relatively small group of children. We did not detect any association between performance on the Babyscreen app and previous touchscreen use. Touchscreen use in our population is extremely common from a young age. The effect of socioeconomic factors, maternal education and touchscreen exposure will require study of larger cohorts from a variety of backgrounds.

CONCLUSION

We have shown that a 10–15 min touchscreen tool, which is language and administrator independent, can be used to screen for children at risk of cognitive delay at 18–24 months.

ORCID iDs
Neil Marlow http://orcid.org/0000-0001-5890-2953
Deirdre M Murray http://orcid.org/0000-0002-2201-9912

REFERENCES

1 Bayley N. *Bayley scales of infant and toddler development: Bayley-III: Harcourt assessment, Psych Corporation*, 2006.
2 Aylward GP. Continuing issues with the Bayley-III: where to go from here. *J Dev Behav Pediatr* 2013;34:697–701.
3 Cristia A, Seidl A. Parental reports on touch screen use in early childhood. *PLoS One* 2015;10:e0128338.
4 Twomey DM, Wrigley C, Ahearne C, *et al*. Feasibility of using touch screen technology for early cognitive assessment in children. *Arch Dis Child* 2018;103:853–8.
5 Spencer-Smith MM, Spittle AJ, Lee KJ, *et al*. Bayley-III cognitive and language scales in preterm children. *Pediatrics* 2015;135:e1258–65.