A randomised controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder

Running title: Randomized-controlled trial of an iPad-based app for ASD early intervention

A randomised-controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder

Andrew J.O. Whitehouse[1], Joanna Granich[1], Gail Alvares[1], Margherita Busacca[2]; Matthew N. Cooper[1], Alena Dass[1], Thi Duong[3], Rajes Harper,[2] Wendy Marshall[4], Amanda Richdale[5], Tania Rodwell[4], David Trembath[6], Pratibha Vellanki,[3] Dennis W. Moore[2], and Angelika Anderson[2].

[1]Telethon Kids Institute, The University of Western Australia, Perth, Western Australia, Australia

[2]Krongold Centre, Faculty of Education, Monash University, Melbourne, Victoria, Australia.

[3]School of Information Technology, Deakin University, Geelong, Victoria, Australia

[4]Autism West, Perth, Western Australia

[5]Olga Tennison Autism Research Centre, La Trobe University, Melbourne, Australia

[6]Griffith Health Institute, Griffith University, Brisbane, Queensland, Australia

Word count: 6,721

Conflict of Interest: Wendy Marshall and Tania Rodwell were employed by Autism West at the time of this study. Autism West is a not-for-profit organisation that receives royalties for sales of the TOBY app. No author (including Wendy Marshall and Tania Rodwell) receives, or has ever received, personal remuneration from sales of the TOBY app.

Abstract

Background: Technology-based interventions for Autism Spectrum Disorder (ASD) have proliferated, but few have been evaluated within the context of a randomised-controlled trial (RCT). This RCT evaluated the efficacy of one technology-based early intervention program (Therapy Outcomes By You; TOBY) in young children with ASD.

Methods: TOBY is an app-based learning curriculum designed for children and parents as a complement to early behavioural intervention. Eighty children (16 female) were recruited to this RCT within 12 months of receiving a diagnosis of ASD (M age = 3.38; SD = 0.69) and randomised to receive either treatment-as-usual (community-based intervention, n = 39) or the TOBY therapy (at least 20 minutes/day) plus treatment-as-usual (n = 41) for a period of six months. Outcomes were assessed at 3- and 6-months post-baseline. (Australian New Zealand Clinical Trials Registry: ACTRN12614000738628; www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=365463).

Results: Children in the TOBY intervention group averaged 19 minutes/day engaging with the app in the first 3 months, but only 2 minutes/day during the second 3-months. There was no group difference in scores on the primary outcome, the Autism Treatment Evaluation Checklist, at either the 3- or 6-month follow-up. However, significant improvements at the 6-month follow-up were observed in the TOBY intervention group relative to the treatment-as-usual group on three secondary outcomes: the Fine Motor and Visual Reception subscales of the Mullen Scale of Early Learning and the Total Words Understood scale of the MacArthur-Bates Communicative Development Index. Statistical trends towards improvement in the TOBY intervention group were observed on measures of adaptive function, though these decreased in magnitude from the 3- to 6-month follow-up.

Conclusions: This study provides evidence that technology-based interventions may provide a relatively low-cost addition to existing therapist-delivered interventions for children with ASD. However, sustained use of the app over the full six-month period was a challenge for most families.

Keywords: Autism Spectrum Disorder; randomised controlled trial; early intervention, computer assisted learning.

Abbreviations: ATEC: Autism Treatment Evaluation Checklist; BFRS-R: Behaviour Flexibility Rating Scale – Revised; CSBS: Communication and Symbolic Behavior Scales Developmental Profile Caregiver Questionnaire; EIBI: Early and Intensive Behavioral intervention; MCDI: MacCarthur-Bates Communication Development Inventory; MSEL: Mullen Scales of Early Learning; TAU: Treatment as

usual; TOBY: Therapy Outcome By You (app); RBS-R: Repetitive Behavior Scale-Revised; VABS-II: Vineland Adaptive Behavior Scales–Second edition.

A randomised-controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is the collective term for neurodevelopmental disorders characterised by qualitative impairments in social interaction and communication, and a restricted range of activities and interests (American Psychiatric Association, 2013). There is an accumulating evidence base for the benefits of a range of early and intensive behavioural interventions (EIBI) seeking to improve developmental outcomes in ASD (Magiati, Tay, & Howlin, 2012). Systematic reviews (Howlin, Magiati, Charman, & MacLean, 2009; Warren et al., 2011) and meta-analyses (Makrygianni & Reed, 2012) of these interventions identify several key features linked to efficacy, including the early onset of treatment, data driven decision-making, and high intensity delivery. However, despite the emergence of high-quality evidence from randomized controlled trials (RCT) for the efficacy of certain EIBI programs (Dawson et al., 2010), there remain several challenges to their community implementation. Amongst the most significant challenges is that the efficacy of many programs depends on considerable therapy time with trained clinicians. This time requirement, which can range from 20 to 40 hours per week (see Eikeseth, 2009), and associated substantial financial expense (Peters-Scheffer, Didden, Korzilius, & Matson, 2012), means that a significant proportion of families are unable to access these programs. The development of treatment delivery mechanisms that increase access to EIBI while maintaining treatment efficacy will provide considerable benefit to the ASD community.

The use of computer technology has several potential benefits for facilitating broader access to EIBI, including increasing the time exposed to therapy at relatively low cost, and providing an interactive platform that can help structure and guide a home therapy curriculum for caregivers (Ramdoss et al., 2012). Several studies have reported benefits of specifically designed applications (‘apps’) for improving a range of social and emotional skills in children with ASD, including social responsiveness (Beaumont & Sofronoff, 2008; Fletcher-Watson et al., 2015), social conflict resolution (Bernard-Opitz, Sriram, & Nakhoda-Sapuan, 2001), and mental state understanding (Swettenham, 1996). The recent development of touch-screen tablets (e.g., iPad™) has increased both accessibility and affordability, and thus the potential utility of computer technology in ASD intervention, even further.

The Therapy Outcomes By You (TOBY) app is one ASD intervention program that is accessible on a touch-screen device and targets a wide range of developmental abilities. TOBY is not designed to replace individual therapy with clinicians, but rather provide a means for structuring home-based therapy under the guidance of caregivers. The TOBY curriculum targets four major areas of early learning: (1) *Visual and auditory understanding*, which targets the perception and discrimination of basic sensory information (e.g., colour, shape, same-ness, difference); (2) *Imitation*, which asks users to copy an action, design, or pre-speech sound; (3) *Receptive and expressive language*, which requires children to recognise and produce object and action names; and (4) *Social skills*, such as eye gaze, gestures, facial expressions, and joint attention. The TOBY app is designed to facilitate learning through Applied Behaviour Analysis principles, and modifies stimuli, reinforcement and prompting in response to three fundamental task types. *Solo tasks* can be completed by the child on the iPad without assistance from a caregiver (e.g., the child is asked to find a given stimulus picture from a set of pictures). *Partner tasks* require assistance from a caregiver in some way, either to recognise the child's response, such as an action or spoken word, or to present stimuli, prompts, and reinforcement (e.g., the child is presented with a picture of an object on the iPad and is asked by the caregiver, "What is this?"). *Natural environment tasks (NET)* focus on the generalisation of skills learnt during Solo and Partner tasks, and are performed off the iPad, and incorporated into a child's daily routines. For each task type, the child's response is inputted to the iPad (either by the child or by the caregiver), and machine learning algorithms are used to individualise syllabus progression.

In a series of case studies, we have established the acceptability and feasibility of the TOBY app as an addition to EIBI for children with ASD (Moore et al., 2015). The current study evaluated efficacy through a RCT design. Over a period of 6 months, we examined the developmental progress of a group of children with ASD who used the TOBY app in addition to receiving their regular community-based therapy, in comparison to a group of children who received their regular community-based therapy only. We hypothesised that use of the TOBY app as an addition to community–based therapy would help support home-based therapy, and thus facilitate greater improvement in a range of ASD-specific and broader developmental skills. Given the TOBY app has a focus on the core ASD symptoms of communication (*receptive and expressive language*) and social (*social skills*) difficulties, we selected as the primary outcome an ASD-specific assessment designed

to measure change over time in these and other ASD symptoms. Standardised measures of broader developmental and functional ability that are not specific to ASD, but are also targeted by the TOBY syllabus and widely used in ASD research, were selected as secondary outcomes. Tertiary outcomes were additional measurements of the key ASD behavioural domains of communication, social skills and repetitive behaviours.

Methods

Trial design and study settings

This was a multicentre, stratified (site, socio-economic status, and developmental quotient, with 1:1 randomisation), parallel-group RCT conducted at two university-based research centres in Australia: The Telethon Kids Institute, University of Western Australia (Perth, Western Australia) and the Krongold Centre, Monash University (Melbourne, Victoria). The trial was registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12614000738628; www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=365463). Further methodological information, including inclusion/exclusion criteria, detailed description of the intervention and outcome measures, and power analyses, are provided as a supplementary file.

Participants

The final sample comprised 80 children, randomized to either the “TOBY group” (n = 41, 80.6% male) or the “Treatment as Usual” (TAU; n = 39, 76.9% male) group. Modules 1 (55% of participants) or 2 (45%) of the Autism Diagnostic Observation Scale-Generic (ADOS-G; Lord et al., 2000) was also administered at trial entry. Figure 1 provides the CONSORT diagram for the trial, and family and sociodemographic characteristics of the study participants are presented in Supplementary Table 2. The study was approved by the Human Research Ethics Committees of the University of Western Australia (RA/4/1/5656) and Monash University (CF13/991-2013000482),

Procedures

The length of the intervention period was 6-months, with follow-up assessments taking place at the mid-point (3-months post-baseline) and conclusion (6-months post-baseline) of this period. At baseline, 3- and 6-month follow-ups, children received a range of psychometric assessments, administered by researchers blind to study group. Prior to each assessment, questionnaires were mailed to caregivers, who were encouraged to return these at the time of assessment, or by mail up to 14 days after the assessment.

Interventions

Treatment group: TOBY intervention

Caregivers of children randomised to the TOBY intervention group received an initial 2-hour training session conducted by a trained therapist, during which they were familiarised with the TOBY curriculum, and the entry point in the curriculum was determined for each child in consultation with caregivers. Caregivers were then loaned an iPad with the TOBY app installed, and asked to complete at least 20 minutes/day of TOBY-based therapy for the next six months. Time using the TOBY app was measured using the on-iPad recording mechanism, which logged the duration of time the child spent on learning trials. Caregivers were asked to weave TOBY-directed activities into their child's natural environment, and so the in-app recording mechanism may be an underestimate of the total time spent participating in TOBY-directed activities. However, this log facilitates a broad understanding about fidelity to the study requirement of 20 minutes/day using the TOBY app.

Caregivers were contacted by the study team every fortnight during the trial period to encourage use of the TOBY app, to provide an opportunity for caregivers to ask any questions, and for the research team to enquire about perceived barriers to use of the TOBY app. Fortnightly phone calls were scheduled in advance around dates in which the family may be unavailable (e.g., family vacations), and when telephone contact could not be made with families, an email was sent. The study design placed no restrictions on participants' use of other behavioural therapies, but TOBY therapy was not specifically integrated into the community therapy by the study team. Caregivers were asked to complete daily paper-and-pen diaries regarding any community therapy their child had received during the six month trial period. At the conclusion of the therapy period, caregivers were asked to provide written feedback by listing 'three features that (they) liked or disliked about the TOBY app.

Control group: Therapy as Usual (TAU)

The TAU group were not prevented from receiving community-based therapy, but were asked not to download and use the TOBY app, nor any other app designed for ASD therapeutic use. To incentivise adherence to this protocol, families in the TAU group were provided with a loan iPad with the TOBY app installed at the conclusion of the trial. The nature and quantity of any community therapy received was recorded via caregiver-completed home diaries.

Measures

Primary Outcome

The Autism Treatment Evaluation Checklist (ATEC) is a parent-completed checklist that provides a quantitative measure of abilities in a range of developmental skills relevant to ASD (Rimland & Edelson, 2000). Higher scores on each subscale represent greater phenotypic severity.

Secondary Outcomes

The Mullen Scales of Early Learning (MSEL) and Vineland Adaptive Behavior Scales–2nd edition (VABS-II) served as the secondary outcomes for this trial. The MSEL provides an assessment of developmental abilities across five domains. Around one-third (29%) of children in the current study received T-scores on one or more subscales at or below 20, representing 3 or more standard deviations below the mean. As these scores were not considered reliable indicators of developmental abilities, nor likely to be sensitive to change over time, analyses were conducted using the age equivalents (in months) for each subscale (Akshoomoff, 2006). The VABS-II is a parent-report checklist that assesses adaptive behaviour in four domains (Sparrow, Cicchetti, & Balla, 2005). Standard scores are generated for each domain subscale, as well as a Total Adaptive Behaviour Composite, based around a mean of 100 and standard deviation of 15. Lower scores indicate greater impairment in adaptive functioning.

Tertiary Outcomes

The tertiary outcomes were a series of caregiver-report questionnaires. Language and social development were assessed with the "Words and Gestures" form of the MacCarthur-Bates Communication Development Inventory (MCDI; Fenson et al., 1993), and the Communication and Symbolic Behavior Scales Developmental Profile Caregiver Questionnaire (CSBS; Wetherby & Prizant, 2002). For both scales, higher scores indicate greater abilities. Repetitive and restricted behaviours and interests were assessed with the Repetitive Behavior Scale-Revised (RBS-R; Bodfish, Symons, & Lewis, 1998) and the Behaviour Flexibility Rating Scale (BFRS-R; Peters-Scheffer et al., 2008). Higher scores on these scales indicate greater frequency and severity of repetitive behaviours, and greater behavioural inflexibility, respectively.

Statistical analyses

Participants were analysed on an intention-to-treat basis. The study hypotheses were assessed using an analysis of covariance framework. Baseline assessment score was entered as the independent variable in a linear regression model, and the 3- or 6-month assessment as the outcome

variables. Visual checks for a linear relationship between variables and homoscedasticity of residuals were used to confirm model fit requirements were met. Beta coefficients and 95% confidence intervals (95%CI) for the between group differences between baseline and relevant follow-up assessment are presented. Statistical significance was defined as p <.05, but results where p <.1 are also reported.

As a supplementary analysis, we used linear mixed effects models to examined change in test scores between the 3- and 6-month assessments. This analytical approach conforms to intention-to-treat principles as it allows for the inclusion of all participants into analyses, regardless of missing data, and is the recommended approach for evaluating psychiatric clinical trials (Hamer & Simpson, 2009). We investigated the coefficient for the time variable following adjustment for baseline scores, using an autoregressive covariance matrix to define within-subject error (coefficients estimated via maximum likelihood). Interaction terms (group X time) were fitted to assess difference in change between groups across the 3- and 6-month assessments. Models fitted with and without the interaction term were compared via a likelihood ratio test. Linear mixed effects modelling was carried out using the nlme R package (Pinheiro, Bates, DebRoy, & Sarkar, 2016), and all models were adjusted for study site.

Results

Baseline characteristics

Five families withdrew from the study after eligibility determination and randomization, but before completing all baseline assessments, and so were subsequently removed from further analysis. The remaining 75 participants completed at least one cognitive test and the ATEC at baseline, with missing data across the tests ranging from 1 to 4 data points/person. Baseline characteristics are presented in Table 1. The most common reasons for missing data included behavioural non-compliance by children and failure to return questionnaires by parents. In total, 12 individuals did not take part in the follow-up assessments and were recorded as “discontinued intervention” (see Figure 1). Chi-square analyses and independent-samples t-tests revealed no statistically significant differences between the children who withdrew (n = 12) and the remaining trial participants for sex, group allocation or study site, nor on ADOS-G calibrated severity scores (Gotham, Pickles, & Lord, 2009), MSEL, or VABS-II scores at baseline. At least one cognitive test was completed by 65 participants (32 TOBY group) at the 3-month follow-up (missing data ranging from 10 to 15 data points) and 63 participants (30 TOBY group) at the 6-month follow-up (missing

data ranging from 13 to 21 data points). Further descriptive statistics (including sample sizes for each comparison) are presented in Supplementary Tables 3a-b and 4a-b.

Comparison of baseline test scores between the two sites identified several statistically significant differences on the ATEC speech subscale (difference between means 5.3; 95% CI 1.5, 9.2; $p < 0.01$), MSEL Fine Motor subscale (difference between means 6.3; 95% CI 0.7, 11.8; $p = .03$) and all VABS subscales (difference between means ranging from 6.9; 95% CI 0.5, 13.3; $p = .03$, to 14.1; 95% CI 6.6,21.4; $p < .01$), which varied in direction. All statistical modelling was adjusted for study site.

<Please insert Table 1 about here>

Trial fidelity

On average, children in the TOBY intervention group spent a median of 1593 (IQR 419, 2775) minutes using the TOBY app in the first three months and 23 (IQR 0,555) minutes in the second three months of the trial ($p < .01$). This represents a per day median of 19 (IQR 5,33) and 2 (IQR 0,46) for the first and second 3-months of the trial ($p < .01$), respectively. Thus, while children in the TOBY group were averaging just under the recommended amount of time on the TOBY app (at least 20 minutes/day) in the first three months, there was a large drop in usage in the second three months. There was a substantial range in time spent using the TOBY app (0-3 months: 1-800 minutes/week; 3-6 months: 0-836 minutes/week). Further data on TOBY use are provided in Supplementary Table 5.

There was no difference between the first and second 3-month intervention periods in the number of fortnightly telephone calls made to families in the TOBY intervention group (Supplementary Table 6), The barriers to TOBY use that were reported by caregivers during these phone calls include a lack of time to organise therapy materials and implement the off-iPad activities, relationship difficulties, family illness and significant life-events (Supplementary Table 7). It was also reported that some children lacked interest in the on-iPad tasks, whilst others were hyper-sensitive to the visual and auditory stimuli presented by the TOBY app.

Primary outcome

The adjusted difference in Total ATEC score for the TOBY group relative to the TAU group, was -5.8 units (95%CI; -13.6,2.0, $p = .14$) and 4.4 units (95%CI; -5.5,14.3, $p = .37$) at 3- and 6-months respectively after adjustment for baseline scores (Table 2). However, neither result reached statistical significance. No significant differences were observed between groups for any of the four ATEC subscales at either the 3 or 6-month assessments, though the 3-month Communication subscale

showed a trend towards greater improvement in the TOBY group, -2.1 units (95%CI; -4.5,0.3, p = .08). Significant interaction terms (p <.05) for treatment by time (3- to 6-month score change) were observed for the ATEC total score and the Communication subscale (Supplementary Table 3a), which represents symptom improvement over time in the TAU group and a lack of improvement in the TOBY group.

<Please insert Table 2 about here>

Secondary outcomes

Participant scores on the MSEL and VABS-II across the trial are presented in Table 2. After adjustment for baseline scores, a pattern of group differences were observed at the 3- and 6-month assessments for the Visual Reception and the Fine Motor skills subscales of the MSEL. For Visual Reception this was quantified as 4.0 units (95% CI; -0.5,7.5; p = .03) and 4.5 units (95% CI; -0.1,8.9; p = .05) and for Fine Motor skills 2.9 units (95% CI; -0.2,6.0; p = .07) and 5.0 units (95% CI; 0.9,9.1; p = .02) at the 3- and 6-month assessment, respectively. No differences were observed for the Receptive or Expressive Language subscales of the MSEL. (MSEL T scores are provided as Supplementary Table 8). Analysis of VABS-II scores at the 3-month assessment identified adjusted differences between groups for the Total Adaptive Behavior Composite and the Daily Living and Motor Skills subscales. The largest difference, and only difference significant at the p<.05 level, was observed for the Daily Living Skills subscale at 6.5 units (95% CI; 0.6,12.3, p = .03). At the 6-month follow-up, this difference was of a lesser magnitude and did not reach statistical significance (6.3 units; 95% CI -2.8,10.0; p = .26). No significant interactions terms for the change between the 3- and 6-month assessments were observed for the MSEL or VABS-II scales (Supplementary Table 3a). A graphical presentation of the change in primary and secondary outcomes is provided as Supplementary Figure 1.

Tertiary outcomes

A significant baseline-adjusted difference was observed between groups on the MCDI Total Words Understood scale, 27.3 units (95% CI; 4.4,50.3; p = .02) and 26.3 units (95% CI; 6.9,45.6; p = .01) at the 3- and 6-month assessments, respectively (Table 3). No further group differences on the MCDI were observed at the 3- or 6-month assessment. There were no group differences on the RBS Total Score, and the majority of BFRS subscales. The one exception was the BRFS Subscale of “Flexibility towards the environment” at the 6- months assessment, for which the TOBY group had a

significantly reduced score compared to the TAU group, -2.3 units (95% CI; -4.4,-0.2; p = .03). No interaction term for the change between the 3- and 6-month assessments on the tertiary outcomes reached statistical significance (Supplementary Table 4a).

Testing potential benefits of apps

In addition to evaluating the efficacy of the TOBY app, the study sought to test whether the use of apps may increase the time a child is exposed to therapy, and also whether iPads may help structure and guide a home therapy curriculum for caregivers. Data from parent diaries indicated no significant (p = .80) difference between groups in the amount of community intervention received in terms of both the average number of minutes per week (TAU: n = 30, M = 284.7, SD = 270.8; TOBY: n = 27, M = 271.9, SD = 300.85) nor the number of therapy sessions per week (TOBY: n = 22, M = 2.65, SD= 1.56; TAU: n = 27, M = 2.91, SD = 2.10). Week-by-week data for community therapy received are presented in Supplementary Table 9. When the time using the TOBY app was added to the total amount of therapy received, the TOBY group were found to receive more therapy (in minutes) per week than the TAU group (TOBY: n = 27, M = 361.6, SD = 321.0; TAU: n = 30, M = 284.7, SD = 270.8), though this difference did not reach statistical significance (p = .31; Supplementary Figure 2).

At the conclusion of the 6-month follow-up assessment, caregivers of children in the TOBY intervention group were asked to list up to three features that they liked or disliked about the TOBY app. The caregivers of 24 participants listed 62 (65.3%) “like” statements and 32 (34.7%) “dislike” statements (Supplementary Table 10). The most frequent “like” statement, provided by n = 17 (70.8%) caregivers, related to TOBY providing a helpful therapy planning tool, with new ideas for therapy and activities. Other common statements were that TOBY was easy to use (n = 10, 41.7%) and that the app provided a positive learning experience for their child (n = 9, 41.7%) with an attractive structure and layout (n = 9, 41.7%). The most common “dislike” statement was that the off iPad activities were too time consuming to prepare (n = 10, 41.7%).

Discussion

The current RCT investigated the efficacy of the TOBY app in improving developmental abilities in recently diagnosed children with ASD. Across the 6-month period, there was no difference between the two groups on the primary outcome of autistic symptoms. However, analyses of secondary and tertiary outcomes identified that the TOBY group had relatively greater improvement than the TAU group on a range of developmental skills, including visual reception and fine motor skills

(MSEL), and single word comprehension (MCDI). There was also a trend at the 3-month assessment for the TOBY group to have relatively greater improvement on functional abilities measured by the VABS (Total Composite and Daily Living Skills), though the between-group differences did not achieve statistical significance. The large number of between-groups comparisons warrant caution in the interpretation of the study findings, as does the higher discontinuation rate in the TOBY group (23.1%) compared to the TAU group (8.3%). However, the broad trends observed across multiple measures indicate that the use of an app that guides caregivers in home-based therapy may be a useful complement to existing community therapy in promoting a select range of developmental abilities.

This study was motivated by two potential benefits of incorporating app-based therapies into EIBI for ASD. First, app technology may increase the amount of time a child is exposed to therapy, at a comparatively cheaper cost (i.e., the cost of the iPad and app) than therapist-delivered intervention. This benefit was observed in the current study; the two groups received a similar amount of community-based therapy (though, the quality and nature of this therapy could not be compared between groups), but the use of the app as part of this study increased the total amount of therapy received by the TOBY group. While TOBY-directed therapy is likely to differ in quality to therapist-delivered intervention, machine learning algorithms incorporated into the app allows treatment to be progressively tailored to a child based on previous responses to learning trials. Second, iPad apps may provide a therapy-delivery platform that can help structure and guide a home therapy curriculum for caregivers. This benefit was also broadly observed in the current study, with the majority of caregivers providing feedback at the conclusion of the trial (70.8%) that the TOBY app provided a helpful therapy planning tool.

However, while these two benefits were validated by the current study, the fidelity data raise an important methodological consideration for future trials of iPad-based therapy apps intended to be used outside of clinical settings. Despite considerable effort from the research team to ensure that participants in the TOBY group used the app for 20 minutes per day – participants were telephoned every fortnight to encourage app use – the majority of participants did not reach this target for the full six month therapy period. One possibility is that the TOBY app was insufficiently motivating for children, though we note that caregivers provided broadly positive feedback regarding their child's interest in the app (Supplementary Table 10), and only one child discontinued intervention due to a

lack of interest (Figure 1). In the current trial, reasons documented by parents for the reduced time spent using the app spanned family, parental, and child factors (Supplementary Table 7), many of which are also well-established barriers to more traditional home-based interventions (Johnson & Hastings, 2002). The attrition rate was higher in the TOBY compared to the TAU group, and similar barriers to therapy were reported by the study participants who withdrew from the trial and did not complete the follow-up assessments. Among the individuals who completed the trial, app use declined over the course of the therapy period. It is possible that the broader and larger between-groups effects observed at the 3-month follow-up compared to the 6-month follow-up on several of the VABS-II subscales are a consequence of the decline in app use across the trial period. However, we were unable to adequately test this hypothesis given concerns that the in-app time log is likely to underestimate the total time the child was engaged in TOBY-directed activities off the iPad.

The pattern of findings observed in the current trial – greater improvement in the treatment group on select secondary outcomes measuring developmental skills, but not on primary outcomes measuring ASD symptomatology – has also been observed in RCTs of a number of EIBI programs (Dawson et al., 2010; Green at al., 2010). The current lack of suitable measures of core ASD symptoms for use in clinical trials is a subject of considerable discussion (Fletcher-Watson & McConachie, 2015). Current “gold standard” diagnostic measures, such as the ADOS (Lord et al., 2000), were not designed to assess change over time, and thus may be insensitive to small improvements generated by interventions over the short term. The primary outcome in the current trial, the ATEC, was developed specifically to measure incremental improvement in ASD symptoms, and there is evidence of good internal consistency, and adequate content and predictive validity (Magiati, Moss, Yatges, Charman, & Howlin, 2011). However, the ATEC has been criticised for incorporating items assessing both developmental ability and ASD symptom severity in each subscale, which makes it difficult to isolate and identify symptom-related change over time (Charman, Howlin, Berry, & Prince, 2004). Furthermore, in the current study, the ATEC was completed by parents, who were unblinded to their child’s treatment group. Despite these ongoing challenges for ASD intervention research, we can conclude from the current trial that use of the TOBY app over a period of six months did not lead to a reduction in ASD symptom severity.

Parts of the TOBY syllabus require the interactive use of the touch-screen device to complete therapeutic activities, and so it is unsurprising that we observed significant improvements amongst

this group in fine motor skills. While we did not quantify time the TAU group spent using an iPad during the trial period, we were able to confirm that no child in this group used an app designed for therapy for children with ASD. The fine motor improvement in the TOBY group relative to the TAU group may represent a therapeutic effect of touch-screen device use, and may be an important component for other therapy apps to consider. The Visual Reception scale of the MSEL assesses visual perceptual skills such as matching and sorting of objects and pictures across a range of categories (colours, animals, size), as well as memory for these items. Several modules within the TOBY app target these skills, suggesting that the significant improvement in the intervention group is a direct therapeutic effect of using the TOBY app. While there was no difference between groups in the level of improvement on the Receptive and Expressive Language subscales of the MSEL across the trial period, the TOBY group did have significantly greater improvement on a caregiver-report measure of receptive vocabulary (MCDI). Substantial research has found caregiver report to be a valid measure of early vocabulary development, particularly for children with developmental disorders, in ascertaining a more representative measure of abilities that may not be observed during a direct assessment (Luyster, Kadlec, Carter., & Tager-Flusberg, 2008). However, in the current study, caregivers were not blind to group allocation, which raises the risk that the caregiver-report measures may have overestimated language improvement in the TOBY intervention group. This same reporting bias may also be relevant to the small improvements in functional ability observed in the TOBY relative to TAU groups at the 3-month follow-up measured using the caregiver-report VABS-II.

In conclusion, the current study provides evidence that an iPad-based app that assists in structuring and delivering home-based therapy for children with ASD can lead to small improvements in several developmental skills, but not lessen autistic symptom severity. Another key outcome of the study was that the use of apps to guide home therapy do not circumvent the well-established barriers to therapy fidelity in this setting. Despite these challenges, the current findings provide encouragement that well-designed therapeutic apps can provide assistance to caregivers in delivering therapy, and increase time a child is exposed to therapy at relatively low cost. We emphasise that the current study protocol included elements that may not be available in the typical clinical setting, such as the provision of iPads and fortnightly contact with families to encourage therapy fidelity. Future trials that focus both on the generalisability of these findings and how barriers to app use in the family home may be circumvented, will be able to build on the findings from the current study.

Key points

- Early intervention can be beneficial for children with ASD, but many families are unable to meet the financial and time requirements of these therapy programs.
- The development of relatively low-cost computer software applications (apps) provides one possible means of complementing early intervention.
- This study randomised 80 young children with ASD to receive either community therapy as usual, or community therapy as usual plus 20 minutes/day using the therapy app.
- While there was no difference between the two groups in autistic symptom severity after 3- or 6-months of app use, there were broad trends for improvement among the children engaging with the therapy app in fine motor and visual reception skills, and single-word comprehension.
- Therapy apps may provide an efficacious complement to traditional early intervention.

Acknowledgments

We acknowledge support from the Helen MacPherson Trust, as well as a private donation from an anonymous donor (through the Australian Children's Trust). Andrew Whitehouse is support by a Senior Research Fellowship from the National Health and Medical Research Council (APP1077966), and David Trembath is funded by an Early Career Research Fellowship from the National Health and Medical research Fellowship (GNT1071811). These funders had no role in the study design, analysis or drafting/submission of the manuscript. We would also like to express our thanks to the participants for taking part in the study.

Correspondence to:

Professor Andrew Whitehouse

Telethon Kids Institute

University of Western Australia

100 Roberts Rd, Subiaco

Western Australia 6008

Email: Andrew.Whitehouse@telethonkids.org.au

Phone: +61 8 9489 7770

Facsimile: +61 8 9489 7700

References

Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *12*(4-5), 269–277.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Washington, DC: APA Press.

Beaumont, R., Sofronoff, K. (2008). A multi-component social skills intervention for children with Asperger Syndrome: The Junior Detective Training Program. *The Journal of Child Psychology and Psychiatry, 49*,743–753.

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders*, *31*, 377–384.

Bodfish, J.W., Symons, F., & Lewis, M. (1998). *The repetitive behaviour scale: Test manual.* Morganton: Western California Center Research Reports.

Charman, T., Howlin, P., Berry, B., & Prince, E. (2004). Measuring developmental progress of children with Autism Spectrum Disorder on school entry using parent report. *Autism, 8*(1), 89-100.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., et al. (2010). Randomized, controlled trial of an intervention for toddlers with Autism: The Early Start Denver Model. *Pediatrics, 125*(1), e17-e23.

Eikeseth, S. (2009). Outcome of comprehensive psycho-educational interventions for young children with autism. *Research in Developmental Disabilities, 30*(1), 158-178.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., et al. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore: Paul H. Brookes Publishing Co

Fletcher-Watson, S., & McConachie, H. (2015). The search for an early intervention outcome measurement tool in autism. *Focus on Autism and Other Developmental Disabilities, 32,* 71-80.

Fletcher-Watson, S., Petrou, A., Scott-Barrett, J., Dicks, P., Graham, C., O'Hare, A. et al. (2016). A trial of an iPad™ intervention targeting social communication skills in children with autism. *Autism*, *20*(7), 771-782.

Goin-Kochel, R.P., Myers, B.J., Hendricks, D.R., Carr, S. E., & Wiley, S.B. (2007). Early responsiveness to intensive behavioural intervention predicts outcomes among preschool children with autism. *International Journal of Disability, Development and Education*, *54*(2), 151-175

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders. 39*(5):693-705.

Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., et al. (2010). Parent-mediated communication-focused treatment for pre-school children with autism (MRC PACT); a randomized trial. *The Lancet, 375, 2152–2160.*

Hamer, R.M., & Simpson , P.M. (2009). Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials. *American Journal of Psychiatry, 166*(6), 639-641.

Howlin, P., Magiati, I., Charman, T., & MacLean, J.W.E. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities, 114*(1), 23-41.

Johnson, E., & Hastings, R.P. (2002). Facilitating factors and barriers to the implementation of intensive home-based behavioural intervention for young children with autism. *Child: Care, Health and Development, 28*(2), 123-129.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(8), 1426-1438.

Magiati, I., Moss, J., Yatges, R., Charman, T., & Howlin, P. (2011). Is the Autism Treatment Evaluation Checklist a useful tool for monitoring progress in children with autism spectrum disorders? *Journal of Intellectual Disability Research, 55*(3), 302–312.

Magiati, I., Tay, X.W., & Howlin, P. (2012). Early comprehensive behaviorally based interventions for children with autism spectrum disorders: A summary of findings from recent reviews and meta-analyses. *Neuropsychiatry, 2*(6), 543-570.

Makrygianni, M.K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 4*, 577–593.

Moore, D. W., Venkatesh, S., Anderson, A., Greenhill, S., Phung, D., Duong, T., et al. (2015). TOBY play-pad application to teach children with ASD. *Developmental Neurorehabilitation, 18*(4), 213-217.

Mullen, E. (1995). *The Mullen scales of early learning*. Circle Pines, MN: American Guidance Service.

Peters-Scheffer, N., Didden, R., Korzilius, H., & Matson, J. (2012). Cost comparison of early intensive behavioral intervention and treatment as usual for children with autism spectrum disorder in the Netherlands. *Research in Developmental Disabilities, 33*(6), 1763-1772.

Pinheiro. J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2016). _nlme: Linear and Nonlinear Mixed Effects Models_. R package version.

Peters-Scheffer, N., Didden, R., Green, V. A., Sigafoos, J., Korzilius, H., Pituch, K. … Lancioni, G. (2008). The behavior flexibility rating scale-revised (BFRS-R): Factor analysis, internal consistency, inter-rater and intra-rater reliability, and convergent validity. *Research in Developmental Disabilities, 29*(5), 398-407.

Ramdoss, S., Machalicek, W., Rispoli, M., Mulloy, A., Lang, R., & O'Reilly, M. (2012). Computer-based interventions to improve social and emotional skills in individuals with autism spectrum disorders: A systematic review. *Developmental Neurorehabilitation, 15*(2), 119-135.

Rimland, B., & Edelson S. (2000). *Autism Treatment Evaluation Checklist: Statistical Analyses*, San Diego, CA: Autism Research Institute.

Sparrow, S.S., Cicchetti, D., & Balla, D.A. (2005). *Vineland Adaptive Behavior Scales-2nd edition manual*. Minneapolis: NCS Pearson Inc.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry*, 37(2), 157–165.

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics, 127*(5), e1303-e1311.

Wetherby, A., & Prizant, B. (2002). *CSBS DP Manual: Communication and Symbolic Behavior Scales Developmental Profile*. Baltimore, MD:Paul H Brookes Publishing Co.

Table 1. Characteristic and baseline measures for the treatment as usual (TAU) and TOBY intervention groups. Between groups comparisons are conducted with independent samples t-tests.

	TAU (*n* = 36)	TOBY (*n* = 39)	*p* value
Age at diagnosis (months)	35.26 (8.8)	34.45 (8.8)	.70
Age at baseline (months)	40.25 (8.41)	39.36 (8.50)	.65
ADOS Calibrated Severity Score	5.94 (2.12)	6.00 (2.14)	.89
ATEC Total	74.08 (22.67)	68.85 (24.17)	.33
MSEL age equivalents, months			
Receptive language	24.72 (13.36)	24.82 (15.71)	.97
Expressive language	25.64 (12.56)	26.44 (15.67)	.80
Visual reception	30.11 (15.16)	28.49 (14.19)	.63
Fine motor	28.31 (11.30)	27.90 (10.88)	.87
VABS, standard scores			
Total Composite [a]	71.27 (11.59)	77.69 (16.18)	.06
Communication [b]	72.31 (14.60)	76.00 (18.58)	.35
Socialization [b]	70.54 (10.46)	74.87 (13.60)	.14
Daily Living Skills [b]	77.43 (12.89)	83.92 (19.98)	.11
Motor Skills [b]	79.64 (13.16)	85.53 (15.54)	.10

[a] TAU *n*=33, TOBY *n*=36
[b] TAU *n*=35, TOBY *n*=38

Table 2. Primary (ATEC) and Secondary (MSEL and VABS) outcomes as change from baseline, by treatment group and follow-up assessment.

	3-month outcome			6-month outcome		
	TAU [a]	TOBY [a]	Baseline adjusted group difference [b]	TAU [a]	TOBY [a]	Baseline adjusted group difference [b]
ATEC [c]						
Total	-10.12 (15.35)	-14.72 (14.64)	-5.81 (-13.61,2.00;0.14)	-15.72 (19.25)	-9.85 (15.66)	4.43 (-5.46,14.31;0.37)
Communication	-2.50 (4.46)	-4.10 (5.09)	-2.11 (-4.50,0.27;0.08)	-5.76 (6.15)	-2.58 (7.30)	2.26 (-1.11,5.63;0.18)
Sociability	-3.25 (6.26)	-3.41 (4.29)	-0.92 (-3.64,1.79;0.50)	-4.48 (5.58)	-2.27 (5.03)	1.53 (-1.29,4.36;0.28)
Sensory	-1.03 (4.76)	-2.41 (5.52)	-1.71 (-4.35,0.93;0.20)	-3.24 (5.91)	-1.58 (7.70)	0.75 (-2.77,4.27;0.67)
Physical	-3.34 (7.78)	-4.79 (9.14)	-2.10 (-5.96,1.75;0.28)	-2.24 (9.25)	-3.42 (8.16)	-1.87 (-6.44,2.70;0.41)
MSEL[d]						
Receptive language	3.94 (6.92)	6.03 (8.08)	1.89 (-1.86,5.64;0.32)	8.53 (8.06)	8.54 (9.31)	-0.37 (-4.89,4.15;0.87)
Expressive language	3.06 (4.94)	2.87 (5.64)	-0.18 (-2.87,2.51;0.89)	5.19 (7.58)	3.82 (10.09)	-1.38 (-6.08,3.31;0.56)
Visual reception	3.00 (7.26)	7.16 (6.79)	4.00 (0.49,7.52;0.03)	5.73 (6.91)	10.39 (10.19)	4.50 (0.08,8.91;0.05)
Fine motor	2.21 (5.94)	5.16 (6.72)	2.90 (-0.21,6.02;0.07)	3.88 (6.86)	8.93 (9.09)	5.01 (0.91,9.11;0.02)
VABS-II [d]						
Total Composite	0.50 (6.03)	4.23 (9.70)	3.96 (-0.51,8.42;0.08)	2.62 (7.79)	5.18 (10.50)	2.20 (-2.93,7.32;0.39)
Communication	1.41 (8.62)	5.47 (12.15)	4.46 (-0.89,9.81;0.10)	4.16 (8.94)	7.70 (11.22)	4.10 (-1.01,9.20;0.11)
Socialization	0.62 (8.05)	1.81 (13.24)	2.59 (-2.90,8.08;0.35)	4.65 (7.67)	3.90 (13.24)	0.01 (-5.65,5.67;1.00)
Daily Living Skills	-0.06 (9.99)	3.97 (13.77)	6.45 (0.60,12.30;0.03)	2.71 (10.61)	4.83 (13.96)	3.62 (-2.76,10.00;0.26)
Motor Skills	-0.67 (8.70)	2.03 (10.66)	4.63 (-0.65,9.92;0.08)	0.69 (9.96)	1.54 (11.44)	2.55 (-3.32,8.42;0.39)

[a] Mean (SD) for change (difference between follow-up and baseline assessment)
[b] Beta coefficient (95%CI, p value) for the TOBY group relative to the TAU group, from a linear regression model adjusted for baseline measurement
[c] Higher scores indicate greater level of difficulty
[d] Lower scores indicate greater level of difficulty

Table 3. Tertiary outcomes presented as change from baseline, by treatment group and follow-up assessment.

	3-month outcome			6-month outcome		
	TAU [a]	TOBY [a]	Baseline adjusted group difference [b]	TAU [a]	TOBY [a]	Baseline adjusted group difference [b]
MCDI (Words and Gestures) [d]						
Total words produced	24.60 (50.18)	29.29 (77.76)	5.84 (-28.74,40.41;0.74)	56.39 (60.60)	56.04 (49.89)	1.42 (-29.75,32.58;0.93)
Total words understood	6.57 (37.47)	33.93 (48.34)	27.32 (4.36,50.29;0.02)	21.64 (29.55)	47.00 (42.22)	26.25 (6.94,45.55;0.01)
Total gestures	1.79 (7.06)	3.54 (6.64)	1.06 (-2.61,4.73;0.57)	4.68 (6.10)	5.88 (7.61)	1.84 (-2.05,5.74;0.35)
CSBS [d]						
Social Composite	1.53 (2.94)	2.53 (4.00)	1.11 (-0.72,2.93;0.23)	2.07 (3.88)	3.65 (3.46)	1.71 (-0.31,3.73;0.09)
Speech Composite	1.17 (2.46)	0.63 (2.66)	-0.44 (-1.66,0.79;0.48)	1.04 (3.66)	0.77 (2.98)	-0.08 (-1.73,1.56;0.92)
Symbolic Composite	1.63 (2.51)	1.27 (2.50)	0.07 (-1.14,1.27;0.91)	1.32 (4.02)	1.00 (2.51)	0.23 (-1.44,1.90;0.78)
RBS-R [c]						
Total Score	-4.97 (11.07)	-4.03 (13.89)	0.35 (-5.36,6.05;0.90)	-2.71 (19.56)	-5.09 (17.61)	-2.61 (-13.01,7.80;0.62)
BFRS-R [c]						
Total Score	0.13 (4.97)	-1.30 (6.56)	-1.48 (-4.59,1.62;0.34)	0.10 (6.46)	-1.85 (6.56)	-2.73 (-5.84,0.38;0.08)
Flexibility towards objects	0.17 (1.70)	0.00 (2.51)	-0.20 (-1.30,0.89;0.71)	-0.03 (2.04)	-0.30 (2.23)	-0.50 (-1.53,0.52;0.33)
Flexibility towards environment	-0.17 (3.47)	-1.33 (3.81)	-1.23 (-3.17,0.72;0.21)	0.34 (4.32)	-1.48 (4.45)	-2.31 (-4.42,-0.20;0.03)
Flexibility towards people	0.13 (1.91)	0.04 (1.63)	-0.17 (-1.03,0.69;0.70)	-0.21 (1.61)	-0.07 (1.62)	-0.02 (-0.71,0.67;0.95)

[a] Mean (SD) for change (difference between follow-up and baseline assessment)
[b] Beta coefficient (95%CI, p value) for the TOBY group relative to the TAU group, from a linear regression model adjusted for baseline measurement
[c] Higher scores indicate greater level of difficulty
[d] Lower scores indicate greater level of difficulty

Figure captions

Figure 1. CONSORT flow diagram showing study enrolment, randomization, follow-up and analysis.