Using Tablets to Collect Data from Young Children

Michael C. Frank, Elise Sugarman, Alexandra C. Horowitz,

Molly L. Lewis, & Daniel Yurovsky

Department of Psychology, Stanford University

Abstract

Mobile, touch-screen devices are increasingly ubiquitous in children's lives. The extensive use of such devices provides both a challenge and an opportunity for developmental psychologists. While more research is needed to understand the nature and consequences of young children's interactions with such devices, they nevertheless present an exciting opportunity for data collection. We describe a simple method for creating cross-platform, interactive tablet experiments using open web-based resources. We illustrate this method by collecting data from children 1–5 years old in a simple word-recognition paradigm and show that it yields reliable accuracy and reaction time data; we provide code for this experiment for the use of other researchers. Tablets should be considered as a viable method for collecting low-cost, well-controlled developmental data.

Introduction

Since Apple debuted the iPad in 2010, the ubiquity of personal tablets—interactive, touchscreen-based computing devices that are typically larger than a phone but smaller than a laptop—has skyrocketed. By 2013, an estimated 75% of families with young children in the United States owned a mobile device, with 40% of households owning tablets (20% of lower-income families and 63% of higher-income families; ?, ?); tablets were estimated to constitute 19% of children's overall screen time from ages 0–8. Accordingly, the market for downloadable applications targeted to children has increased dramatically, representing 72% of the top selling apps in the iTunes store, and over 80% in the Education category (?, ?). The explosion in popularity of touchscreen devices has created a surge of interest in both the consequences of tablet use for children and the potential for using these devices as scientific tools.

Nevertheless, there are significant challenges involved in using tablets for research. These include the difficulty and potential costliness of developing custom applications ("apps") for individual experiments, technical problems with app distribution and cross-device compatibility, and the lack of data confirming the reliability of data collected with tablets. Perhaps for these reasons, there are currently few published articles that have used tablets to collect developmental data.

The goal of the current article is to address these challenges. We begin by briefly reviewing research on the use of tablets by children. We then describe a method for creating tablet-based experiments using standard, freely-available web-development tools. Although this method requires some programming experience, it is significantly easier than developing tablet-native applications (e.g., apps that can be downloaded from the Apple App Store or Google Play platforms and are specific to a particular device) and has relatively few drawbacks from the research perspective. We show how this method can be used to implement a simple word-recognition experiment and report data from that

experiment. Our data suggest that the tablet platform is an engaging method for data collection that, when used appropriately, can yield reliable accuracy and reaction time from young children.

Prior Work on Children's Tablet Use

The primary focus of research on tablets with children has been evaluation of their educational potential. We briefly review this literature and then turn to the issue of characterizing children's interactions. The starting point for investigations of learning from tablets is the suggestion that children learn less from passive video viewing than they do from live demonstrations, a phenomenon known as the "video deficit" effect (e.g. ?, ?, ?, ?). These findings influenced the American Academy of Pediatrics (AAP) to uphold a policy recommending no "screen time" for children under age 2 (?, ?).

The experiences offered by touch-screen devices may be distinct from passive viewing of television and computer screens, however. ? (?) notes that the AAP's published media recommendation of no screen time was made before the release of tablet devices, and suggests that the policy should be revised to distinguish and accommodate some time with touch-screens. He proposes that tablets can offer a number of potential benefits that many traditional toys do not. Specifically, tablets are reactive (responding contingently to the childs actions), interactive (eliciting responses from the child based on his or her actions), tailorable (modifiable according to age or preference), progressive (can increase in difficulty or complexity with experience), highly portable, and able to promote joint attention (a child and caregiver can interact with the program together).

Interactive technologies may pose a number of benefits over other passively-viewed screen alternatives. ? (?) found that 24- to 30-month-olds successfully learned a novel verb from a live interaction via video chat, while a yoked group who saw the same input with non-contingent presentation failed to learn. ? (?) found that 15-month-olds were

able to object transfer knowledge learned via a touch-screen to a physical toy. Practice with interactive games may also benefit attention and motor control (?, ?), and more generally screen time with tablets can promote independence, opportunities for physical closeness, and feelings of fun and accomplishment (?, ?). The contingent feedback and individualization offered by mobile devices may thus make them a better medium for learning and engagement than traditional view-only screens. Such benefits may also make them valuable tools for research and evaluation.

One concern with children's performance using tablets is the accuracy of their motor control. Experience likely plays a substantial role in modulating the accuracy of interactions, especially for the youngest children, and some gestures are easier than others. ? (?) tested 2- to 5-year-olds in Malaysia and the United Kingdom on their ability to perform the seven most common touch gestures of iPads: tap, drag/slide, free rotate, drag and drop, pinch, spread, and flick. By age 2, children mastered the tap and drag/slide gestures, and by age 3, they could reliably produce all by the spread gesture. By age 4, all children performed all of the gestures. ? (?) additionally suggest that enabling multi-touch responding—allowing touches from more than one location to be recorded simultaneously—may reduce motor control issues of children trying to interact with the tablet while holding or resting their hand on other parts of the screen.

In sum, research on tablets suggests that interactive media may provide more learning opportunities than passive viewing. Tablets are extremely engaging to children and may provide a valuable source of information, but designers of tablet materials for young children should take care to consider appropriate content and gesture capabilities for their target age group. We next turn to the advantages of tablets for research purposes.

Why Collect Developmental Data With Tablets?

Researchers working with young children have access to a wide variety of experimental paradigms, ranging from behavioral forced-choice to eye-tracking and looking time-based measures (?, ?, ?). Nevertheless, each of these methods has its drawbacks. Behavioral methods are flexible and easy to implement, but they can be subject to experimenter biases, and often require labor-intensive offline coding. In contrast, corneal-reflection eye-tracking methods are precise and unbiased, allowing tight experimental control, but can be expensive and are difficult to implement outside of tightly-controlled lab settings. In addition, eye-tracking paradigms are typically very limited in length as they tend to be non-interactive (cf. ?, ?), and often result in substantial data loss due to excessive movement, calibration issues, etc.

Tablet-based experimental paradigms can supplement both of these data collection modalities by allowing researchers to create engaging, interactive paradigms that are low cost but nevertheless offer some of the precision and experimental control of eye-tracking paradigms. In the sections below, we offer five reasons why tablets may be valuable tools for developmental research.

Diverse measures. Tablets allow the collection of many different dependent variables, including forced-choice responding and gesture trajectories. We especially highlight the ability to measure reaction times using tablets. Reaction times provide a continuous, informative measure of cognitive process and have proven invaluable in both developmental (e.g. ?, ?, ?) and adult (e.g. ?, ? et seq.) studies. Behavioral paradigms that afford reaction time measurement for young children typically require time-consuming offline coding. In contrast, eye-tracking paradigms yield reaction times, but with the tradeoff that the paradigms typically have limited interactivity.

Engagement. Children across a wide range of ages enjoy tablets and are highly motivated to explore new tablet-based tasks, allowing researchers to collect more data from individual children. As documented by ? (?) and many informal observations of infants' and young children's interest, the use of tablet-based devices may lead children to view tablet-based experiments as similar to engaging commercial apps. In addition, behavioral paradigms with young children often require a "warm-up" period so that the child is comfortable interacting with an unfamiliar experimenter; such a period can be shorter or absent to the extent that a tablet-based paradigm requires less face-to-face interaction.

Experimental control. Studies that involve interaction with an experimenter are the de-facto standard in cognitive development research, but such studies have as major drawbacks that it may be difficult or impossible to ensure that experimenters are blind to condition and hypothesis. Often hypotheses are simple enough that even "naïve" research assistants can infer the experimental structure or the desired response. While there are methods for avoiding such issues, the possibility of eliminating experimenter bias via computerized stimulus presentation is an attractive method for situations where it is possible. And even when bias is not introduced, there are significant challenges in training a group of research assistants to implement an experimental protocol uniformly. Thus, tablet-based experiments can alleviate concerns about bias and uniformity of experimental presentation.

Inclusion of special populations. Tablets can be a valuable tool for working with children who have learning or physical disabilities (e.g. ?, ?, ?, ?, ?). The lack of face-to-face interaction demands make tablets especially appealing for working with children on the autism spectrum, who may find interactions with an unfamiliar experimenter especially aversive.

Scale. The low cost and high accessibility make it possible to distribute tablets to research assistants who can collect data in a wider variety of locations (though not without concern about environmental distractions). In addition, it is in principle possible that children could contribute data to experimental studies via experiments administered by their parents or caregivers.[1] It is this latter possibility that might be truly revolutionary in terms of the scope of data collection, though there are many challenges that this sort of data collection method would need to overcome.

Despite all of these advantages, there are significant technical issues related to app creation that have slowed the adoption of tablet-based methods in developmental research. Creating tablet-native apps (that can be accessed through an Android or iOS tablet's main menu) requires substantial programming expertise—and worse yet, this expertise is not transferrable across tablet platforms. Developing a native app for iOS requires programming in Objective-C using Apple's proprietary app development tools; development for Android requires programming in a Java variant. Both languages have specific software development tools (SDKs) that allow programmers to tailor their apps to particular devices but that require extensive time commitment for mastery.

Although there are cases where having a robust, native app might lead to better performance or adoption, the cost of developing such apps is significant. An experiment that requires extensive media or animation, use of the tablet's camera, or large-scale data transfer is likely to perform better as a natively written app. Similarly, standardized experimental tools are more easily distributed and used via Google Play or the Apple App Store, the standard distribution methods for commercial software. To create such apps, though, either an investigator must learn a special-purpose programming language, or a programmer must be employed. This latter option is more common, but costs can easily

---

[1]For an example of this method, see `http://lookit.mit.edu`, a web-based platform for looking-time experiments that engages parents as experimental confederates.
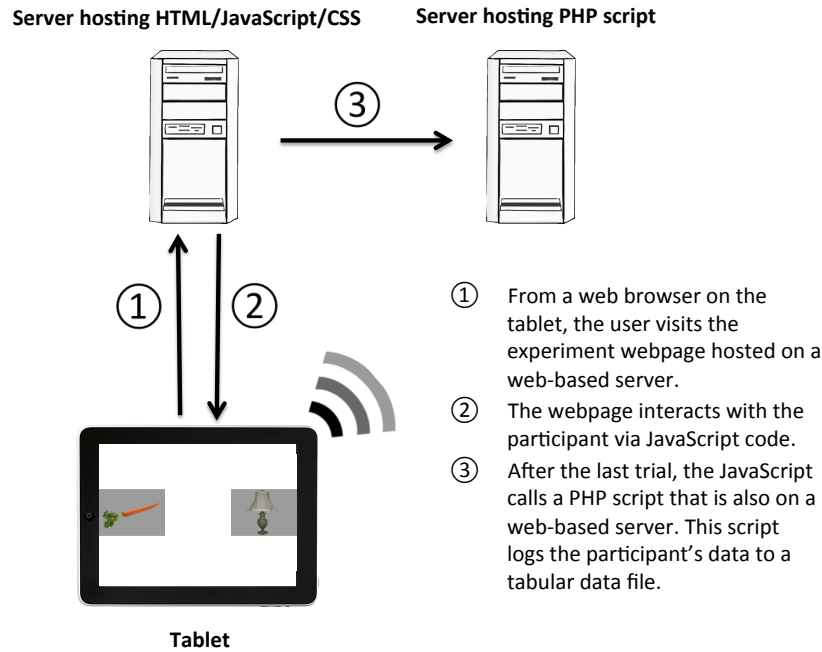
Figure 1. A schematic of our method of collecting tablet data using a web-based experiment. Note that the two server functions can be executed by the same system; we have separated them here for clarity.

run in the tens of thousands of dollars; in this sort of scenario, making scientifically-useful changes to an app may be undesirable because of added costs.

A Web-Based Method for Developing Tablet Experiments

In contrast to costly, special-purpose methods, here we describe an alternative method for prototyping and collecting data on tablets. The method relies on creating the experiment as a simple webpage and then showing this page in the tablet's web browser. This method still requires some programming experience, but much less. In addition, it draws on open-source, freely-available tools that are relatively easy to master compared with native app development tools. Experimental "apps" made via this method are easily

shared, modified, and ported across platforms, leading to faster iteration and data collection. In addition, our method offers the ability to collect control data with adult participants on Amazon Mechanical Turk (?, ?, ?) with virtually no changes to the underlying experiment. Our method has three components (see Figure **??**):

1. A HTML/JavaScript/CSS web page, which is the core of the experiment,

2. A server-side PHP script to collect data in tabular format, and

3. A tablet device running a kiosk application to present the experiment.

We discuss each below.[2]

Web-based experiment. The first component of our system is a web-based experiment created using HTML, JavaScript, and CSS. This combination is currently the most common set of tools for the creation of interactive websites. HTML (HyperText Markup Language) forms the framework for the static content shown on the website, while CSS (Cascading Style Sheets) gives a consistent set of formatting options. JavaScript then controls the interaction functionality of the website. In particular, we make use of a JavaScript library called jQuery (`http://jquery.com`) that allows HTML elements like images, layouts, text, and media to be dynamically added, subtracted, and rearranged as participants interact with the resulting page.

It is beyond the scope of this article to provide an introduction to creating JavaScript-based web experiments, but there are a wealth of available resources on this topic. These resources include both psychology-specific tutorials and more general introductions to the general HTML/JavaScript/CSS model. A sample experiment (as well as the server side code) is provided in the version control repository for this paper at `http://github.com/langcog/tablet/` and can be viewed directly at

---

[2]Our work has used the Apple iPad as the primary tablet platform and we provide details that are tailored for this platform. Nevertheless, and unlike other methods, our method is in principle portable to a wide range of mobile devices with only very minor modifications.

`http://langcog.stanford.edu/tablet/tabletcentered.html`.

Hosting a web-based experiment requires server space, which is often provided by universities but can also be purchased from commercial providers. We find that the requirements on such a server are typically not high, so the costs involved are minimal.

Server-side data collection. The second component of our system is a simple script that allows the experiment to save data in tabular format. Although native apps can store information on tablets, browser-based apps such as ours cannot; thus, our method requires that the data be saved to a separate server. Every time a participant completes the experiment, their data is sent to this script, written in PHP (a simple scripting language that is often used for backend web development) and hosted on a server with script execution capabilities. The script simply appends data from the experiment to a tabular data file. The script can be hosted on any web-accessible server that has been configured to run executable files of this type. In practice, this will often be the same server that hosts the experiment itself, though it need not be. Data are then retrieved by the research from this server using SFTP (Secure File Transfer Protocol) or a similar method.

Tablet configuration. The third component of our system is an internet-connected tablet. The requirement for internet connectivity is perhaps the most substantial drawback of our method, relative to native apps. The tablet must either have wireless access (e.g., from the testing location or a mobile hotspot device) or direct cellular connectivity. Once the tablet has access to the web, the tablet's browser can simply be directed to the experiment website.

A major challenge of presenting experiments in the tablet's web browser is to ensure that children cannot accidentally or intentionally navigate away from the experiment, exit the browser, or change perspective (e.g. by zooming). In practice, we use two tools for this purpose: the first is Guided Access, a mode on the iPad that disables the hardware

buttons. The second is Mobile Kiosk, an app that further locks the iPad into a particular view of a webpage. The combination ensures that the tablet presents the experiment as intended. With both of these systems in place, the "look" of the experiment is extremely similar to a native app.

An Experimental Test of the Reliability of Tablet Data

To test the reliability and broad developmental applicability of our tablet method, we implemented a simple word-recognition paradigm based on prior eye-tracking work (?, ?, ?, ?). We tested this paradigm with a cross-sectional sample of children ages 1–5 from a local children's museum. Our paradigm both tested children's recognition of familiar words and also asked them to make "mutual exclusivity" (ME) inferences—that a novel label refers to a novel word rather than a familiar competitor (?, ?). We hypothesized that familiar word recognition would be easy for all but the very youngest children in our sample, while the mutual exclusivity inference would be more challenging for all children.

Methods

Participants. We recruited a sample of 86 children 1–5 years old from the floor of Children's Discovery Museum of San Jose. Demographics for these participants are shown in Table **??**. An additional 18 children in the target age range were recruited but excluded from the sample for the following reasons: parent-reported English exposure less than our pre-specified criterion of 75% (7), parent interference (4), issues with the iPad pre-training (4), experimenter error or technical issues (2), and self-reported developmental disability (1).

Stimuli. Visual stimuli consisted of images of sixteen familiar and eight novel objects. All were presented as cropped images on a gray background with approximately the same resolution and detail. Audio stimuli consisted of a carrier phrase ("Can you find

Table 1

Demographic information for our participants.

Age group	N	Mean age (years)	Proportion male
1-year-olds	18	1.71	0.44
2-year-olds	22	2.46	0.68
3-year-olds	24	3.56	0.67
4-year-olds	22	4.51	0.59

the [target]?"), into which we spliced recordings of the target words. All recordings featured a female speaker and were created following the procedures described in ? (?). We selected a variety of words from the MacArthur-Bates Communicative Development Inventory word list (?, ?, ?), choosing some words that we were certain nearly all children would be very familiar with (e.g., "car," "dog," "bottle") and others that were intended to be slightly more challenging (e.g., "monkey," "shovel," "lion"). Novel words were mono- and di-syllabic words that had been used in previous experiments of this type (e.g. "dax," "pifo," "kreeb"). We further included four fillers to break up the experiment (e.g., a picture of a child rowing a boat with a sentence of audio describing it).

Procedure. We created a simple page to help children learn to tap appropriately on the iPad. To advance past this training page, they needed to tap a set of five dots in random locations; each dot transformed into an "X" when tapped; a second page showed the same game with smiley faces. We found that this fun training sequence helped children select pictures effectively during the rest of the experiment.

Experimental stimuli were then presented in one of two pseudorandom lists, in which trials (pairs of pictures) were counterbalanced for order and target side. The three

Table 2

Proportion children finishing the study and average number of trials out of 28 completed for all children, by age group.

Age group	Proportion finishing	Average number of trials
1-year-olds	0.44	20.4
2-year-olds	0.91	27.5
3-year-olds	1.00	28.0
4-year-olds	0.86	27.6

trial types were (1) *Familiar Word* trials, in which two familiar pictures were presented and children were asked to select the image that matched a familiar word; (2) *Mutual Exclusivity (ME) Control* trials, in which a familiar and a novel image were presented and children were asked to select the image that matched a familiar word; and (3) *ME Inference* trials, in which a familiar and a novel image were presented and children were asked to select the image that matched a novel word. The experiment had 28 trials in total (8 of each of the three trial types plus four fillers).

Results

Children were overall very enthusiastic about the experiment. One-year-olds varied in their ability to complete the study, with most completing half or two thirds but not the full 28 trials. In contrast, the other groups successfully completed the experiment at high rates. Completion rate and average number of trials completed are given in Table **??**.

Accuracies for familiar word trials were above chance even for the one-year-old age group and approached ceiling performance in the two-year-olds (Figure **??**). There were
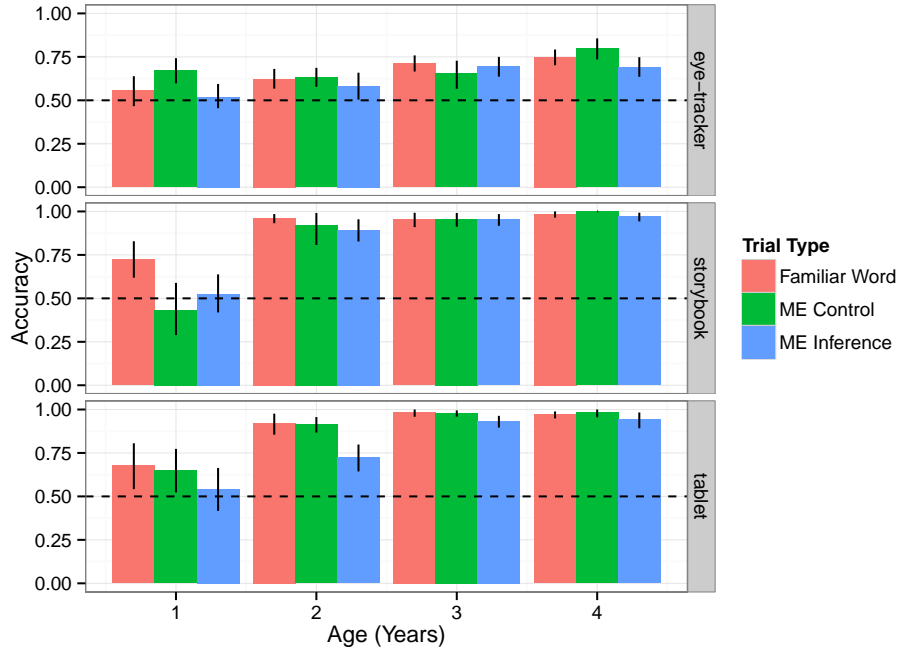
Figure 2. Accuracy in our word recognition task, plotted by age group and trial type. Dashed line shows chance performance. Error bars show 95% confidence intervals, computed by non-parametric bootstrap.

also no substantive differences in accuracy between Familiar Word and ME Control trials. In contrast, ME Inference trials were substantially more difficult for children, with one-year-olds performing at chance, and two-year-olds showing a substantial decrement in performance for these trials.[3] Our data thus support the appropriateness of the

---

[3] To test the reliability of these differences, we fit a logistic mixed effects model to the data (?, ?); Our model specification included age group (coded as a discrete factor with one-year-olds as the intercept), trial type, and interactions of age group and trial type. Models with larger random effect structures did not converge, so we included only a random intercept for each participant. We found a significant increase in performance for all three older age groups ($\beta = 1.68$, 3.21, 2.60, all ps $< .0005$) and a significant decrement in performance for ME Inference trials ($\beta = -1.14$, $p = .0003$), but no effect in ME Control trials and no interactions between trial type and age group (all ps $> .18$).
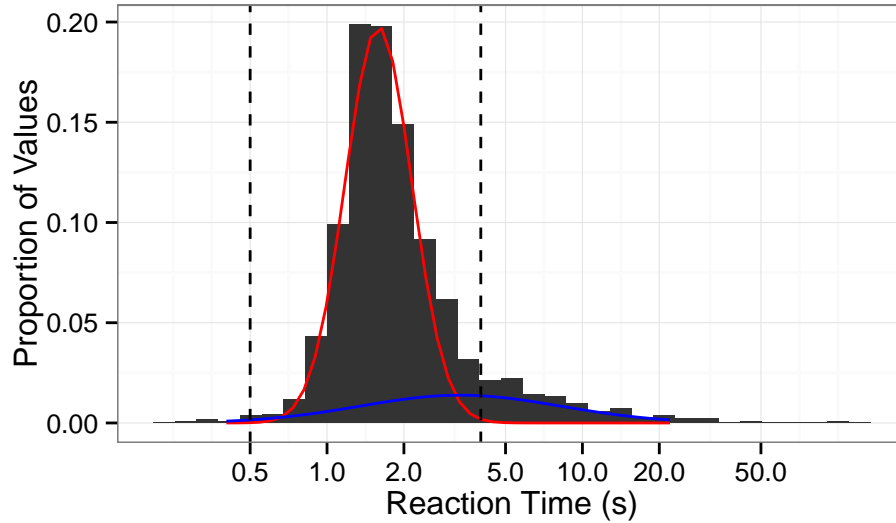
Figure 3. A histogram of reaction times across all participants. Red and blue curves show the best-fitting distributions found by a 2-component Gaussian mixture model, scaled for comparison to the histogram. Dashed lines show the cutoff values we adopted (0.5 and 4.0 s).

tablet-based method for children two years old and above: These children were able to complete our task successfully and performed quite well, on average. As discussed below, accuracies for the one-year-olds (who were mostly above 18 months) were more mixed, but still had some within-subject reliability.

We next turned to the reaction time data. Considering only reaction times for correct trials, we found a substantial skew in the reaction times we collected (Figure **??**). This pattern appeared to be due to two clusters of reaction times: one larger, tighter cluster early on, representing immediate responding, and one broader slower cluster representing trials on which participants were unsure how to respond, looked to a parent, or otherwise deliberated for a substantial amount of time. In an exploratory analysis, we fit a Gaussian mixture model (using the `mclust` package in `R`) and found that the best
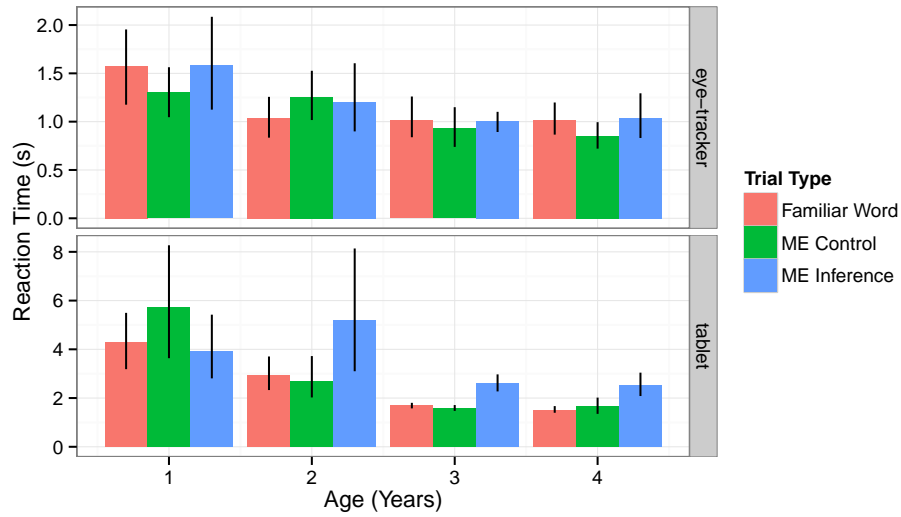
Figure 4. Reaction times in our word recognition task, plotted by age group and trial type. Error bars show 95% confidence intervals, computed by non-parametric bootstrap.

BIC values were indeed produced by a two-cluster solution, which is shown in Figure **??**.

For further analysis, we excluded reaction times faster than 500 ms or slower than 4000 ms. These values were determined post-hoc, based on visual inspection of the distribution. They correspond well both to our intuitions about the task and to the properties of the first mixture component found by the mixture model (for which mean + 3 standard deviations was 3928 ms). Excluding based on these threshold values resulted in a loss of 10.9% of values.[4] In future work, we recommend the adoption of *a priori* standards for reaction time exclusion; we did not do so here because establishing such standards was a primary goal of the current experiment.

Older children were faster overall in our task, and ME inference trials were slower than both Familiar Word and ME Control trials for all groups except the one-year-olds

---

[4]Although this proportion seems high, consider the collection of reaction times based on eye-tracking methods. In these analyses, only half of trials are considered—those where the eyes are on the distractor item at the point of disambiguation (?, ?).

Table 3

Mean reliability values found by a bootstrap sample of random splits of the data from familiar word trials. Values marked SB are computed by the Spearman-Brown prediction formula.

Age group	Accuracy	Reaction time	Accuracy (SB)	Reaction time (SB)
1-year-olds	0.45	0.09	0.61	0.05
2-year-olds	0.31	0.34	0.42	0.44
3-year-olds	-0.12	0.49	-0.27	0.64
4-year-olds	0.03	0.60	0.03	0.74

(Figure **??**). We again did not observe large numerical differences between Familiar Word and ME Control trials.[5] Experimenting with other reaction-time outlier exclusion processes did not qualitatively alter the pattern of results. Nevertheless, we recommend that investigators using this method carefully examine their distribution of reaction times, as averages can be heavily skewed by just a small number of very long RTs.

Group-level developmental differences such as those described above do not necessarily signal a measure with high reliability. We computed split-half reliability coefficients for accuracy and reaction time based on splitting all trials using familiar words (both Familiar Word and ME Control trials) into two halves and computing the

---

[5]A linear mixed effects model fit to log-transformed reaction-time data showed a significant effect of trial number such that reaction times decrease over the course of the experiment ($\beta = .001$, $p = .04$), and three- and four-year-olds were significantly faster than one-year-olds ($\beta = -.19$, $.26$, $ps < .005$). There were no main effects of trial type, but there were reliable slowdowns for ME Inference trials across the two-, three-, and four-year-olds ($\beta = .25$, $.26$, $.29$, all $ps < .0001$). ME control trials were also reliably faster for the four-year-olds ($\beta = -.13$, $p = .04$).

correlation between these two halves. To ensure the robustness of the statistic, we repeated this procedure with 1000 bootstrap replicates (different splits of the trials) and took the average value across these replicates. We also adjusted reliabilities using the Spearman-Brown prediction formula (?, ?) to estimate the overall reliability of the full sample of 16 items.

Reliabilities are shown in Table **??**. Accuracy and reaction time show opposite patterns: while accuracy was somewhat reliable for the youngest two age groups, the presence of a ceiling effect effectively eliminates any reliability for accuracies in the older two groups. In contrast, reaction times were not reliable for the one-year-olds but increased in reliability for the older age groups. These data thus suggest that even a sample of 16 trials can yield reliable data, especially for reaction times in the older age groups. Looking towards future work, the Spearman-Brown formula allows us to compute that a larger sample of trials (e.g. 32 familiar word trials) could potentially yield reliabilities above .65 for two-year-olds and up to .85 for older children. Overall, these data support the use of tablet experiments for the measurement of individual differences, albeit with careful selection of the appropriate measures and experiment length.

Conclusions and Methodological Recommendations

We presented a simple method for creating tablet "apps" that allows experimenters to create basic websites and use these websites to collect data. This method is substantially easier and more flexible than tablet-native app development, which can be costly and time-consuming. In addition, our method allows cross-platform compatibility and easy sharing for the resulting experiments. We showed in a simple word recognition experiment that this method yields reliable reaction time and accuracy data for young children. Data collection using tablets thus has the potential to supplement—and perhaps even supplant—many behavioral methods currently in use.

We summarize our methodological recommendations for the use of tablets:

1. Web-based experiment development can be a valuable tool for designing low-cost tablet experiments. This method should be considered for research applications that can be conducted in a setting with internet connectivity.

2. Tablets are a viable data collection even for very young children (potentially as young as 18 months), but if children have not interacted with a tablet before, care must be taken to familiarize children with the device prior to data collection.

3. Tablets can yield reliable and precise reaction time data, but care must be taken in the analysis to establish a priori procedures for the detection and removal of outliers; otherwise, analyses will be subject to post-hoc bias based on data cleaning choices.

Mobile and tablet-based computers are quickly becoming a ubiquitous part of life—and of childhood. Children play with their parents' and their own devices, and these devices are used for education, reading, games, and video chat with friends and relatives. Many important developmental issues are raised by this ubiquity, and more research must be done to determine how, when, and what children can learn from interactive media. Nevertheless, the opportunity posed by such devices for developmental psychologists is immense.