Tablet assessment of word comprehension reveals coarse word representations in 18–20-month-old toddlers

Chang Huan Lo¹ | Audun Rosslund² | Jun Ho Chai¹ |
Julien Mayor³ | Natalia Kartushina²

¹School of Psychology, University of Nottingham Malaysia, Semenyih, Malaysia

²Department of Psychology & Center for Multilingualism in Society across the Lifespan (MultiLing), University of Oslo, Oslo, Norway

³Department of Psychology, University of Oslo, Oslo, Norway

Correspondence

Audun Rosslund, Center for Multilingualism in Society across the Lifespan, University of Oslo, PO Box 1102 Blindern, 0317 Oslo, Norway.
Email: audun.rosslund@iln.uio.no

Funding information

This work was supported by the following Malaysian grants awarded to J.M.: MOSTI-06-02-12-SF0351. The work of A.R. and N.K. was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223265.

Abstract

The present study explores the viability of using tablets in assessing early word comprehension by means of a two-alternative forced-choice task. Forty-nine 18–20-month-old Norwegian toddlers performed a touch-based word recognition task, in which they were prompted to identify the labeled target out of two displayed items on a touch-screen tablet. In each trial, the distractor item was either semantically related (e.g., dog–cat) or unrelated (e.g., dog–airplane) to the target. Our results show that toddlers as young as 18 months can engage meaningfully with a tablet-based assessment, with minimal verbal instruction and child–administrator interaction. Toddlers performed better in the semantically unrelated condition than in the related condition, suggesting that their word representations are still semantically coarse at this age. Furthermore, parental reports of comprehension, using the Norwegian version of the MacArthur–Bates Communicative Development Inventories, predicted toddlers' performance, with parent–child agreement stronger in the semantically unrelated condition, indicating that parents declare a word to be known by their child if it is understood at a coarse representational level. This study provides among the earliest evidence that

Chang Huan Lo and Audun Rosslund share first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Infancy* published by Wiley Periodicals LLC on behalf of International Congress of Infant Studies

remote data collection in 18-20 month-old toddlers is viable, as comparable results were observed from both in-laboratory and online administration of the touchscreen recognition task.

1 | INTRODUCTION

Historically, studies of early language development involved observations of children's spontaneous speech while interacting with their parents or an experimenter/clinician (Clark, 1974). Despite this method's undeniable appeal of ecological validity, the process of collecting, transcribing, and analyzing spontaneous language samples is labor-intensive and time-consuming.

To go beyond these limitations, researchers have turned to indirect assessment methods—parental reports—that provide insights into their child's communicative–linguistic development. Parental reports systematically utilize parents' extensive experience with their children, and thus allow for the collection of data that is not just more extensive than what can typically be collected during a brief laboratory or clinical session, but that might also be more representative of children's abilities (Fenson et al., 2000). Furthermore, the application of parental reports, such as the widely-used MacArthur–Bates Communicative Development Inventories (CDIs) in cross-linguistic studies, has provided invaluable insight into infants' and toddlers' early lexical development (Bleses et al., 2008; Braginsky et al., 2019; Frank et al., 2021), while other studies have evinced predictive relationships between early vocabulary and subsequent academic outcomes (e.g., Bleses et al., 2016; Duff et al., 2015; Morgan et al., 2015).

Yet, concerns have been raised regarding the sole use of parental reports, in particular when it comes to the assessment of comprehension (more than for production), since parents can at best infer comprehension based on infants' and toddlers' non-verbal responses to language (Feldman et al., 2000; Houston-Price et al., 2007; Tomasello & Mervis, 1994). For instance, while, on a “general level,” previous studies have found moderate to strong correlations between average parental reports on the CDI and direct measures of infants' and toddlers' word knowledge (Fernald & Marchman, 2012; Fernald et al., 2006; Friend et al., 2012; Hurtado et al., 2008), on an “item-level,” the evidence is mixed. For example, studies using indirect, eye-tracking measures revealed both underestimation (Houston-Price et al., 2007) and alignment (Styles & Plunkett, 2009; Syrnyk & Meints, 2017) between parental reports and child comprehension operationalized as visual gaze preference; studies using direct measures (i.e., child's overt answer, a touch response), on the other hand, for example, Friend et al. (2012) and Friend and Zesiger (2011), reported moderate item-level agreement.

Inconsistencies between parental reports and direct measures of child word comprehension might reflect immaturity of children's early lexical-semantic representations, which makes it challenging for parents to pin down whether a child knows a given word. Previous research has shown that early word representations are (semantically) coarse and infants and toddlers use a number of cues to disambiguate words, rather than a one-to-one word–object mapping. For instance, at 6 months of age, infants fail to disambiguate semantically/functionally related items (Bergelson & Aslin, 2017a), and at 8 months, they struggle to disambiguate items matched for frequency in child-directed speech (Kartushina & Mayor, 2019). Although word–object mappings undergo a progressive development through learning, and semantic specificity sharpens by 18–20 months of age (Bergelson & Aslin, 2017b), early word representations remain fragile by the end of the second year (Arias-Trejo & Plunkett, 2010). Arias-Trejo and Plunkett have shown that 18–24-month-olds failed to disambiguate items that were both perceptually and semantically related (e.g., an apple and an

orange), as compared to semantically related but perceptually dissimilar items only (e.g., an apple and a banana), indicating that the presence of a perceptually similar distractor increases the burden of visual discrimination and feature overlap for semantically related objects.

Imprecision of parental reports may have implications when such instruments are used as measures in research or as a basis for decisions in clinical settings (Yoder et al., 1997). For these reasons, the use of supplemental measures to parental reports is encouraged (Dale et al., 2003; Fenson et al., 1993), and further assessment of their validity is needed.

A direct language measure can serve both as a convergent and a supplemental measure of parental reports. While many structured tests, such as the Peabody Picture Vocabulary Test (Dunn, 2018) and the Expressive Vocabulary Test (Williams, 2018), are available to assess young children's vocabulary knowledge, direct measures that are appropriate for assessing children below two years of age remain scarce, due to the inherent difficulty in maintaining infants' and toddlers' interest and attention (Friend & Keplinger, 2003) as well as behavioral non-compliance (Kaler & Kopp, 1990). Whereas looking-based measures, such as the Intermodal Preferential Looking Paradigm (Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996) and the Looking-while-listening task (Fernald et al., 1998, 2006), have been successfully used with infants as young as 4 months old by eliminating the need for a volitional response (Golinkoff et al., 2013), the passive and repetitive nature of such measures may quickly lead to boredom among older toddlers, thus making an extensive assessment impracticable. The Computerized Comprehension Task (CCT; Friend & Keplinger, 2003), on the other hand, is a reliable and valid touchscreen-based measure designed specifically for assessing comprehension among toddlers between 16 and 24 months of age and has been shown to be effective in maintaining children's attention as well as improving compliance (Friend & Keplinger, 2003, 2008; Friend et al., 2012; Friend & Zesiger, 2011; Hendrickson et al., 2015; Poulin-Dubois et al., 2013).

Following the approach of the CCT—in providing an engaging direct language assessment—the present study explores the viability of tablets in assessing toddlers' word comprehension by means of a word recognition task, with the following three objectives. First, despite tablets and apps being increasingly commonplace among children of all ages, the use of tablet-based assessments has been primarily limited to adults and older children. Given that tablets are easy to operate even for the youngest children and additionally, given children's increasing proficiency with tablets (Abdul Aziz et al., 2014; Marsh et al., 2015), there is a need to examine how such devices can be used most effectively to collect child language data. Neumann et al. (2019), for instance, demonstrated that a tablet-based assessment could provide a valid and reliable measure of early literacy skills, at least among the older children ($n = 45$, $M_{\text{age}} = 4.65$) tested in their study. Twomey et al. (2018) further showed that children as young as 24 months old were able to complete a tablet-based assessment of early cognitive functions.

Second, compared to traditional paper-and-pencil tests, tablet-based assessments provide a testing situation that is more engaging and motivating. While the CCT offers the same advantage, the assessment is typically administered in laboratories, where screens are often mounted on a wall or placed on a desk and thus require full arm movements, which may in turn, lead to fatigue in longer sessions (Frank et al., 2016). In contrast, tablet-based assessments require only minimal motor movements and are much more portable due to the small form factor of tablets.

Third, there is a need to further evaluate the alignment between parental reports and children's word comprehension, and, in particular, to assess whether parental evaluations fit best their toddlers' word recognition in coarse (the semantically unrelated condition) or finer-grained contexts (the semantically related condition). Children vary in the strength of their word knowledge at the item-level and capturing this variability is important for a robust understanding of a child's lexical development.

In order to explore the viability of using a tablet-based measure in assessing early word comprehension and to examine the role of semantic relatedness in early word recognition, the present

study employed a two-alternative forced-choice (2AFC) word recognition paradigm (similar to the CCT) with Norwegian toddlers aged between 18 and 20 months. As the CCT is only available in three languages (i.e., English, Spanish, and French), lexical items were selected from the Norwegian adaptation of the CDI–Words and Gestures (CDI–WG; Simonsen et al., 2014), with varying levels of difficulty (defined based on the normative data). Within each trial, toddlers saw on a screen two images: one representing the lexical target, and the other representing the distractor. In contrast to the CCT, in which only semantically related item pairs were used, the current design examined the role of semantic relatedness on toddlers' performance in the word recognition task, by pairing the lexical target with a distractor belonging to a different semantic category (e.g., a car and a cat) and with another distractor belonging to the same semantic category (e.g., a car and an airplane). It was expected that toddlers, in the current study, would be more accurate in semantically unrelated than related trials. Based on previous work using the CCT (Friend & Keplinger, 2003, 2008), accuracy was also expected to mirror the a priori difficulty levels, with accuracy decreasing with increasing difficulty. Finally, if parental reports are an accurate predictor of toddlers' word knowledge, a positive relationship between parent-reported comprehension and toddler's accuracy in word recognition was expected.

2 | METHOD

2.1 | Participants

Parents of 49 monolingual (>75% exposure) Norwegian toddlers (aged between 18 and 20 months) from the Greater Oslo Region, Norway, were contacted to participate in the current study through social media, leaflets distributed in a kindergarten, postal mailing lists, and email lists. After consenting to participate in the study, parents completed the Norwegian adaptation of the CDI–WG (Simonsen et al., 2014) online within one week prior to the study so that the current estimates of their child's vocabulary size could be obtained. Parents' socioeconomic status (SES), indicated by mother's highest education level, ranged from 0 (primary school) to 5 (doctoral degree), with the mean score 3.57 ($SD = 0.82$).

All recruited toddlers were full-term at birth, had no hearing or visual impairments, and had Norwegian as their native language. Toddlers participated in the study in one of three settings: the BabyLing laboratory, a municipal kindergarten, and online (i.e., at toddlers' own homes).¹ In both the laboratory and the kindergarten settings, toddlers were tested by an experimenter, whereas online, toddlers were tested by their parents.² Thus, for simplicity, both the laboratory and kindergarten samples ($n = 21$; 16 females, 5 males) were categorized under the *laboratory* setting, and the online sample ($n = 28$; 15 females, 13 males), the *online* setting. An additional 11 participants had to be excluded for failing to complete the task ($n = 7$; 2 laboratory and 5 online) and for attempting the task more than once ($n = 4$; all online). Mean age, age range, and standard deviation for each setting are detailed in Table 1.

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or a guardian for each child before any assessment or data collection. The study was approved by the ethics committee at the Department of Psychology, University of Oslo and by the Norwegian Centre for Research Data (NSD, ref. 807456).

¹Data were initially collected in the laboratory and kindergarten. Due to the COVID-19 pandemic-related lockdown in Norway (Klesty & Fouche, 2020), data collection proceeded online.

²Parents consented to not to interfere with the task or influence their child's responses.

TABLE 1 Age mean, standard deviation, and range for laboratory and online settings

Setting	<i>M</i> _{age} (months)	<i>SD</i> _{age} (months)	Range _{age} (months)
Laboratory	19.29	0.60	17.91–20.30
Online	19.63	0.63	18.60–20.60

2.2 | Design

The present study used a within-subjects design. Toddlers’ comprehension of 24 lexical items of three levels of difficulty (easy, moderately difficult, and difficult; see Lexical Items section, below) was assessed using a tablet-based 2AFC word recognition task. Lexical targets were assessed under two conditions: semantically related (i.e., the lexical target was presented with a distractor from the same semantic category) and semantically unrelated (i.e., the lexical target was presented with a distractor from a different semantic category).

2.3 | Apparatus and materials

The study was conducted via a custom-based online experimental platform developed by Lo et al. (2021). In the laboratory setting, a Samsung Galaxy Tab S4 was used to run the study, whereas in the online setting, parents’ own touchscreen devices were used. The Norwegian adaptation of the CDI–WG (Simonsen et al., 2014) was used as a measure of vocabulary size.

2.3.1 | Lexical items

Four highly familiar lexical items were selected for the familiarization phase: “ball” [ball], “hus” [house], “sko” [shoe], and “tre” [tree]. For the test phase, a total of 24 lexical items were selected. Each lexical target was assessed twice, by pairing its referent with semantically related and unrelated referents as distractors. Item pairs varied in difficulty (defined a priori on the basis of the Norwegian CDI–WG normative data for 20-month-olds; Frank et al., 2017; Simonsen et al., 2014) and were comprised of an equal number of easy (comprehended by more than 80% of the normative sample), moderately difficult (comprehended by 40%–80% of the normative sample), and difficult (comprehended by less than 40% of the normative sample) items. Within each level of difficulty, there was also an equal representation of animate and inanimate referents. The list of item pairs is provided in Table 2.

2.3.2 | Visual and auditory stimuli

To remove potential biases due to familiarity effects (from assessing the same item twice), visual stimuli for the test phase included 48 images of prototypical referents (as reported by 2 adults in a separate stimuli assessment) for the 24 lexical items assessed (i.e., two images for each item). The set of images used can be found in Appendix 1 (see also Appendix 2 for the images used in the familiarization phase). Within each item pair, the side (left or right) on which a referent appeared was counterbalanced. All auditory stimuli were recorded by a female native speaker of Norwegian in

TABLE 2 Item pairs

Difficulty level	Semantically related	Semantically unrelated
Easy	bil [car]—fly [airplane]	hest [horse]—banan [banana]
	eple [apple]—banan [banana]	hund [dog]—fly [airplane]
	hest [horse]—ku [cow]	katt [cat]—bil [car]
	hund [dog]—katt [cat]	ku [cow]—eple [apple]
Moderate	elefant [elephant]—tiger [tiger]	elefant [elephant]—saks [scissors]
	lastebil [truck]—tog [train]	løve [lion]—tog [train]
	saks [scissors]—blyant [pencil]	sjiraff [giraffe]—lastebil [truck]
	sjiraff [giraffe]—løve [lion]	tiger [tiger]—blyant [pencil]
Difficult	elg [moose]—pingvin [penguin]	elg [moose]—pasta [pasta]
	gås [goose]—ugle [owl]	gås [goose]—shorts [shorts]
	pasta [pasta]—sukkertøy [candy]	pingvin [penguin]—sukkertøy [candy]
	shorts [shorts]—glidelås [zipper]	ugle [owl]—glidelås [zipper]

child-directed speech and then processed in Praat (Boersma & Weenink, 2020) to remove noise and equalize intensity across the 24 prompts.

2.4 | Procedure

The study began with an introductory phase, followed by a familiarization phase and a test phase.

2.4.1 | Introductory phase

Before the familiarization phase began, a smiley face was presented at the center of the screen with an introductory audio “Hei! Har du lyst til å spille?” [Hi! Do you want to play?] to attract participants’ attention. In order to proceed to the familiarization phase, the experimenter/parent had to tap on the “Next” button at the bottom-right corner of the screen.

2.4.2 | Familiarization phase

The familiarization phase consisted of four 2AFC trials to (a) ensure that participants understood the context of the task and (b) familiarize them with the tapping paradigm. In each trial, participants were presented with a pair of highly familiar objects (placed on the left and right sides of the screen respectively) and prompted to tap on the referent for the heard lexical target X embedded in the carrier phrase “Kan du trykke på X?” [Can you touch the X?] Tapping was disabled for the first 2000 ms from the onset of the trial to prevent impulsive responses during the audio prompt that lasted between 1500 and 2000 ms. The timeout was 8000 ms (comparable to Friend et al., 2012), to accommodate for considerable individual variation in response times (see Ackermann et al., 2020). As soon as a (touch) response was provided, the next trial was presented.

2.4.3 | Test phase

Before the test phase began, a smiley face was again presented at the center of the screen, accompanied by an audio with an encouraging phrase “Da fortsetter vi!” [Let's continue!] The experimenter/parent had to tap on the “Next” button to begin the test phase.

The test phase consisted of 48 2AFC trials, in which each lexical target was assessed twice (paired with either a semantically related distractor or a semantically unrelated distractor). In each trial (see Figure 1 for a screenshot), participants were presented with an item pair (see Table 2) and prompted to tap on the referent for the heard lexical target *X* (see carrier phrase from the familiarization phase). Each item pair was presented twice so that each item within the pair served as a target and a distractor in an equal number of trials. As with the familiar trials, tapping was disabled for the first 2000 ms of the trial (to prevent participants from providing responses during the audio prompt that lasted between 1500 and 2000 ms), after which participants were given 8000 ms to respond until the subsequent trial was presented. Trials were presented in a random order, with three breaks interspersed throughout the test phase. During each break, a smiley face was presented in the same manner as before, accompanied by one of the following encouraging phrases: (a) “Da fortsetter vi!” [Let's continue!], (b) “Nå går vi videre!” [Now, we move on!], (c) “Da har vi den neste!” [Then, we have the next (one)!], and (d) “Da er du nesten ferdig! Bra!” [You're almost done! Good!] In order to continue with the test, the experimenter/parent had to also tap on the “Next” button at the bottom-right corner of the screen. Upon completion of the test phase, the smiley face was once again presented, accompanied by an audio with the phrase “Nå er du ferdig! Kjempebra!” [Now you're done! Great!].

3 | RESULTS

The results are organized around three central questions. First, potential differences between data collected online and in-laboratory were considered. Second, the influence of semantic relatedness and difficulty of item pairs on toddlers' motivation to produce a response as well as on their performance in the word recognition task were examined. Finally, the convergent relation between toddlers' performance and parental report (CDI–WG) was assessed. In accordance with previous work using the CCT (Friend & Keplinger, 2003; Friend et al., 2012), missing responses (i.e., trials in which the child did not produce a response) were treated as errors of comprehension.³

3.1 | Attempted trials

The number of trials in which a tap response was produced, regardless of whether the response was correct (i.e., tap on target) or incorrect (i.e., tap on distractor), was used as a measure of toddlers' motivation to produce a response during the word recognition task. Results from a Welch's *t*-test indicated that toddlers who were tested online ($M = 44.286$, $SD = 6.359$) and those who were tested in the laboratory ($M = 40.810$, $SD = 7.061$) did not differ significantly in the number of attempted trials; $t(40.601) = -1.779$, $p = .083$ (see Figure 2).

³The analysis of the number of attempted trials (see below) revealed that toddlers produced less tap responses for difficult trials than for easy trials, suggesting that un-answered trials are not random, but predominantly reveal errors of comprehension—in line with similar observations reported in Friend and Keplinger (2003) and Friend et al. (2012).

FIGURE 1 Screenshot of a trial in the test phase

FIGURE 2 Attempted, correct, and incorrect trials across different settings

To assess whether toddlers' motivation (as indexed by whether an attempt to provide a tap response was made) differed across semantic relatedness and difficulty of the trials, a binomial generalised linear mixed-effects model (GLMM) with a logit link function was fitted and analyzed using the `mixed()` function from the *afex* package (Singmann et al., 2020), which relies on the *lme4* package (Bates et al., 2015) for model fitting. The model included semantic relatedness (related, unrelated), difficulty (easy, moderately difficult, and difficult), toddlers' age (in months), and the interaction between semantic relatedness and difficulty as fixed effects, as well as participant and selected object as random intercepts.⁴ Both semantic relatedness (−1: unrelated; 1: related) and difficulty (−1: easy; 1: moderately difficult, difficult) were sum-coded, whereas age was centered on the mean. To determine a model with a parsimonious random effect structure (Matuschek et al., 2017), the forward “best-path” approach, with $\alpha = .20$ as the inclusion criterion, was used to test random slopes for inclusion (Barr

⁴The inclusion of setting (i.e., online vs. lab) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

TABLE 3 GLMM results for attempted trials

	Model summary			Model comparison		
	β	<i>SE</i>	<i>z</i>	χ^2	<i>df</i>	<i>p</i>
Intercept	3.080	0.281	10.956	103.539	1	<.001***
Relatedness	−0.087	0.075	−1.163	1.355	1	.244
Difficulty				8.516	2	.014*
Moderate	−0.057	0.105	−0.542			
Difficult	−0.249	0.103	−2.432			
Age	0.949	0.395	2.402	5.686	1	.017*
Relatedness:Difficulty				1.618	2	.445
Relatedness:Moderate	−0.106	0.105	−1.006			
Relatedness:Difficult	0.116	0.102	1.136			

* $p < .05$.; ** $p < .01$.; *** $p < .001$.

et al., 2013). As none of the random slopes fell below the inclusion criterion, the random-intercepts-only model was retained:

$$\text{Attempted} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 | \text{Participant}) + (1 | \text{Object})$$

The results are detailed in Table 3, with chi-square statistics and p -values obtained using likelihood ratio tests. Follow-up pairwise comparisons, with p -values adjusted using the Tukey method, were conducted using the `pairs()` function in the *emmeans* package (Lenth, 2020).

As shown in Table 3, there were significant main effects of trial difficulty and age, with the number of attempted trials increasing with age. No significant main effect of semantic relatedness was found; neither did semantic relatedness interact with difficulty. Results from the follow-up tests indicated that toddlers attempted significantly more easy than difficult trials ($\beta = 0.556$, $SE = 0.186$, $z = 2.995$, $p = .008$), while no such difference was found between easy and moderately difficult trials ($\beta = 0.363$, $SE = 0.189$, $z = 1.917$, $p = .134$) as well as moderately difficult and difficult trials ($\beta = 0.193$, $SE = 0.176$, $z = 1.096$, $p = .517$; see also Figure 3).

3.2 | Correct trials

Results from a Welch's t -test indicated that there was no statistically significant difference between toddlers who were tested online ($M = 38.286$, $SD = 7.262$) and those who were tested in the laboratory ($M = 34.095$, $SD = 8.717$) in terms of the number of trials in which they correctly identified the target referent; $t(38.508) = -1.787$, $p = .082$ (see Figure 2). To assess whether toddlers' accuracy differed across semantic relatedness and difficulty of the trials, a binomial GLMM with a logit link function was again fitted and analyzed. The model included the same fixed effects as the previous model (i.e., semantic relatedness, difficulty, age, and the interaction between semantic relatedness and difficulty) and the same random intercepts (i.e., participant and selected object), with by-participant adjustments to the slope of difficulty:⁵

⁵The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

FIGURE 3 Proportion of attempted trials across settings by semantic relatedness and difficulty

$$\text{Accuracy} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 + \text{Difficulty} | \text{Participant}) + (1 | \text{Object})$$

The results are detailed in Table 4, with chi-square statistics and p -values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with p -values adjusted using the Tukey method.

As shown in Table 4, there were significant main effects of semantic relatedness, difficulty, and age. Specifically, toddlers responded with higher accuracy in semantically unrelated than related trials. Toddlers' accuracy also increased significantly with age. No interaction was found between semantic relatedness and difficulty, however. Results from the follow-up tests indicated that toddlers were significantly more accurate in easy trials relative to both moderately difficult ($\beta = 0.523$, $SE = 0.183$, $z = 2.861$, $p = .012$) and difficult trials ($\beta = 1.113$, $SE = 0.164$, $z = 6.799$, $p < .001$). Toddlers were also significantly more accurate in moderately difficult than difficult trials ($\beta = 0.590$, $SE = 0.150$, $z = 3.924$, $p < .001$; see also Figure 4).⁶

3.3 | Convergent validity

At the general level, toddlers' receptive vocabulary size, as measured by the CDI-WG, and their overall accuracy in the word recognition task significantly correlated in both unrelated, $r_{(47)} = .631$, $p < .001$ and related trials, $r_{(47)} = .603$, $p < .001$. Partialling out the effect of age further revealed that

⁶A Spearman correlation between toddlers' overall word recognition accuracy and SES revealed no relationship, $\rho = 0.1$, $p = .46$.

TABLE 4 GLMM results for accuracy

	Model summary			Model comparison		
	β	<i>SE</i>	<i>z</i>	χ^2	<i>df</i>	<i>p</i>
Intercept	1.438	0.143	10.038	56.979	1	<.001***
Relatedness	−0.141	0.054	−2.624	6.782	1	.009**
Difficulty				36.405	2	<.001***
Moderate	0.022	0.097	0.229			
Difficult	−0.568	0.085	−6.660			
Age	0.537	0.193	2.779	7.233	1	.007**
Relatedness:Difficulty				3.887	2	.143
Relatedness:Moderate	−0.114	0.076	−1.511			
Relatedness:Difficult	0.127	0.071	1.785			

* $p < .05$.; ** $p < .01$.; *** $p < .001$.

toddlers' receptive vocabulary size accounted for a significant proportion of unique variance in their recognition accuracy, beyond that accounted for by their age in both unrelated, $r_{(46)} = .593$, $p < .001$, $R^2 = .352$ and related trials, $r_{(46)} = .538$, $p < .001$, $R^2 = .289$.

To explore the consistency between toddlers responses and parent-reported comprehension on the test items (i.e., parent–child agreement), item-level agreement was calculated (see Table 5) and a binomial GLMM with a logit link function was fitted. The model included semantic relatedness, difficulty, age, and the interaction between semantic relatedness and difficulty as fixed effects. Both semantic relatedness (−1: unrelated; 1: related) and difficulty (−1: easy; 1: moderately difficult, difficult) were sum-coded, whereas age was centered on the mean. Random intercepts included participant and selected object, with by-participant adjustments to the slopes of semantic relatedness, difficulty, and their interaction:⁷

$$\text{Accuracy} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 + \text{Relatedness} * \text{Difficulty} | \text{Participant}) + (1 | \text{Object})$$

The GLMM results are detailed in Table 6, with chi-square statistics and p -values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with p -values adjusted using the Tukey method.

Overall, as shown in Table 6, there was good item-level agreement between parental reports and toddlers' responses, although this attenuated with increasing item difficulty. Results from the GLMM indicated that semantic relatedness, difficulty, as well as the interaction between semantic relatedness and difficulty (but not age) significantly predicted parent–child agreement (see also Figure 5). The follow-up tests revealed that parent–child agreement was significantly higher in semantically unrelated than related easy trials ($\beta = 0.795$, $SE = 0.299$, $z = 2.662$, $p = .008$), but no significant differences were found across the different semantic conditions in the moderately difficult ($\beta = 0.253$, $SE = 0.169$, $z = 1.495$, $p = .135$) and difficult trials ($\beta = -0.166$, $SE = 0.164$, $z = -1.014$, $p = .311$).

To further examine whether item-pair comprehension status (i.e., whether the target or the distractor label was known or not known by the toddler as indicated by parental responses on the

⁷The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

FIGURE 4 Accuracy by semantic relatedness and difficulty across different settings. *Note.* Dashed line represents chance (.50)

TABLE 5 Item-level agreement between parental report and toddler performance

Difficulty level	Semantically related	Semantically unrelated	Overall
Easy	.781	.827	.804
Moderate	.615	.661	.638
Difficult	.564	.538	.551
Overall	.653	.675	.664

CDI-WG) was an accurate predictor of toddlers' performance in the word recognition task, another binomial GLMM with a logit link function was fitted, with semantic relatedness, difficulty, item-pair comprehension status, age, and the interaction between semantic relatedness and difficulty as fixed effects. Semantic relatedness (−1: unrelated; 1: related), difficulty (−1: easy; 1: moderately difficult, difficult), and item-pair comprehension status (−1: both unknown; 1: both known, target known only, distractor known only) were sum-coded, whereas age was centered on the mean. Random intercepts included participant and selected object, with by-participant adjustments to the slope of difficulty.⁸

Accuracy ~ Relatedness*Difficulty + Pair Comprehension + Age + (1 + Difficulty|Participant) + (1|Object)

⁸The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

TABLE 6 GLMM results for parent–child agreement

	Model summary			Model comparison		
	<i>B</i>	<i>SE</i>	<i>z</i>	χ^2	<i>df</i>	<i>p</i>
Intercept	0.921	0.163	5.663	68.207	1	<.001***
Relatedness	−0.147	0.066	−2.237	5.436	1	.020*
Difficulty				21.564	2	<.001***
Moderate	−0.240	0.168	−1.423			
Difficult	−0.752	0.182	−4.134			
Age	0.074	0.153	0.486	0.218	1	.641
Relatedness:Difficulty				9.994	2	.007**
Relatedness:Moderate	0.020	0.082	0.249			
Relatedness:Difficult	0.230	0.076	3.030			

* $p < .05$.; ** $p < .01$.; *** $p < .001$.

FIGURE 5 Parent–child agreement by semantic relatedness and difficulty. *Note.* Dashed line represents chance (.50)

The results are detailed in Table 7, with chi-square statistics and p -values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with p -values adjusted using the Tukey method.

As shown in Table 7, parent-reported item-pair comprehension was a significant predictor of toddlers' performance, along with semantic relatedness, difficulty, and age. No significant interaction effect between semantic relatedness and difficulty was found. Results from the follow-up tests indicated that toddlers were significantly less accurate when both target and distractor were reported as unknown compared to when both were known ($\beta = -0.628$, $SE = 0.190$, $z = -3.300$, $p = .005$) and when only the target was known ($\beta = -0.769$, $SE = 0.196$, $z = -3.923$, $p < .001$). No significant differences were found in other cases: (a) both known and target known only ($\beta = -0.141$, $SE = 0.195$, $z = -0.725$, $p = .887$); (b) both known and distractor known only ($\beta = -0.284$, $SE = 0.184$, $z = 1.539$, $p = .414$); (c) target known only and distractor known only ($\beta = 0.425$, $SE = 0.205$, $z = 2.070$, $p = .163$); (d) distractor known only and both unknown ($\beta = -0.344$, $SE = 0.186$, $z = 1.846$, $p = .252$; see also Figure 6).

TABLE 7 GLMM results for accuracy (with parent-reported item-pair comprehension as predictor)

	Model summary			Model comparison		
	β	<i>SE</i>	<i>z</i>	χ^2	<i>df</i>	<i>p</i>
Intercept	1.402	0.144	9.749	58.245	1	<.001***
Relatedness	−0.139	0.054	−2.588	6.586	1	.010*
Difficulty				14.702	2	<.001***
Moderate	0.007	0.098	0.068			
Difficult	−0.403	0.107	−3.776			
Pair comprehension				18.108	1	<.001***
Both known	0.193	0.114	1.685			
Target known	0.334	0.125	2.667			
Distractor known	−0.091	0.117	−0.778			
Age	0.511	0.181	2.817	7.428	1	.006**
Relatedness:Difficulty				4.141	2	.126
Relatedness:Moderate	−0.120	0.076	−1.581			
Relatedness:Difficult	0.132	0.072	1.832			

* $p < .05$; ** $p < .01$; *** $p < .001$.

4 | DISCUSSION

In the interest of developing a performance-based measure of comprehension during the second year of life that addresses the need for a convergent and supplemental measure of parental reports, while taking into account young children's non-compliance and limited attention capabilities (as in Friend & Keplinger, 2003), the present study explored the viability of using a tablet-based 2AFC word recognition task in assessing early word comprehension.

Toddlers aged between 18 and 20 months were tested—either in the laboratory setting by an experimenter or online (i.e., at home) by their parents—on their comprehension of 24 lexical items selected from the Norwegian CDI–WG (Simonsen et al., 2014). During the task, toddlers were asked to identify the referent for the lexical target presented alongside a distractor. Target–distractor pairs were manipulated such that each lexical target was paired once with a semantically related distractor and once with a semantically unrelated distractor. Item pairs also varied in three levels of difficulty (defined based on the Norwegian CDI–WG normative data for age-matched children).

Both the analyses on the number of attempted trials (regardless of whether the response was correct or incorrect) as well as the number of trials in which toddlers provided a correct response revealed no significant differences between the online and laboratory samples, suggesting that toddlers were equally motivated to produce a response in the task and that neither setting led to better or poorer performance. This demonstrates that remote infant data collection with fully automatized tasks can be as efficient and reliable as in situ laboratory assessments. High-quality data through remote administration are not only an important enabler during this time of the global COVID-19 pandemic, but also provide a promising avenue for data collection associated with developmental research, with increased speed, lowered cost, and the potential to an improved sample diversity by reaching to a wider socio-demographic background than traditional laboratory-based research (Sheskin et al., 2020).

Overall, in line with Friend and Keplinger (2008), toddlers attempted significantly more easy than difficult trials. Older toddlers also attempted significantly more trials than younger toddlers. Together,

FIGURE 6 Accuracy by parent-reported item-pair comprehension status. *Note.* Dashed line represents chance (.50)

these findings suggest that toddlers were responding non-randomly and bolster the support for the notion that non-responses represent toddlers' true inability to map the lexical target to its referent, rather than their non-compliance or the lack of motivation, while incorrect responses might be taken as evidence of partial word knowledge, and correct responses—robust word knowledge (Hendrickson et al., 2015).

With regard to the accuracy measure, toddlers demonstrated above-chance performance throughout the task. Congruent with previous work (Friend & Keplinger, 2003, 2008), toddlers' performance was consistent with the a priori "cohort-level" difficulty categorization, as their best performance was obtained for easy trials and their worst performance for difficult trials. As would be expected, older toddlers also performed with greater accuracy relative to younger toddlers.

Examining the role of semantic relatedness, it was found that toddlers displayed more robust recognition in semantically unrelated than related trials, suggesting that, and similar to research in younger infants (Bergelson & Aslin, 2017a), semantical relatedness between the target and the distractor triggered competition effects in referent selection. Although there is evidence that early word representations are semantically more specified by 18–20-months of age (Bergelson & Aslin, 2017b), they still might be lacking representational specificity (Arias-Trejo & Plunkett, 2010). In the current study, in addition to semantic relatedness, lower recognition on some related trials could be attributed to the increased burden of visual discrimination and feature overlap (e.g., both goose and owl are birds and have wings, feather, and a beak), as shown with 18–24-month-olds in Arias-Trejo and Plunkett (2010). It is likely that toddlers, upon hearing the lexical target, co-activated related (and thus, competing) word referents, which subsequently interfered with their lexical decision about the target. Such interference has been reported even among older children, between 3 and 9 years of age, as they took longer to provide a correct response in a visual search task when a related distractor was present than when an unrelated distractor was present (Vales & Fisher, 2019).

Comparing between toddlers' recognition accuracy and their receptive vocabulary size as measured by the CDI-WG, a significant and moderate correlation (comparable to that achieved with the CCT; Friend & Keplinger, 2008) was found, evincing acceptable convergent validity of the word recognition task employed in the present study, and also supporting the feasibility of the CDI-WG, as a general proxy of receptive vocabulary.

Consistent with the CCT (Friend et al., 2012; Friend & Zesiger, 2011), there was also good (albeit not perfect) item-level agreement between toddlers' responses and parental reports across both semantic conditions, with easy items having the highest agreement and difficult items having the lowest agreement. The results further indicated that parent–child agreement was significantly higher in semantically unrelated than related trials, although this was only limited to easy items. This discrepancy suggests that parents' inference on their child's word comprehension is not solely based on evidence of their child's true ability to comprehend the word, but rather on the confluence of both evidence of robust word knowledge (i.e., their child's true ability to comprehend the word) and evidence of partial word knowledge (i.e., their child's ability to respond appropriately when cued by the rich context in which the word is heard, or upon recognizing the sound of the word; Friend et al., 2018; Houston-Price et al., 2007; Tomasello & Mervis, 1994). Restating the finding that toddlers were less accurate in semantically related than unrelated trials, a performance-based measure that uses semantically related target–distractor pairs can potentially tap children's strong, rather than weak, word knowledge to supplement parental reports. Nevertheless, parent-reported item-pair comprehension (i.e., whether the target or distractor label was known or not known by the child) was found to be a significant predictor of toddlers recognition accuracy. Specifically, compared to trials where both the target and distractor were reported by parents as “not understood” on the CDI–WG, toddlers were more likely to respond correctly in trials where either the target or both the target and distractor were reported as “understood,” indicating that parents are adequate informants of their child's language abilities.

It is important to note that while the CCT uses a set of carefully selected test items consisting of an equal representation of nouns, verbs, and adjectives, the present study focused on nouns only. Nevertheless, the good item-level agreement between parental reports and their child's performance provides encouraging results. Supplemented with a principled selection of test items (Chai et al., 2020; Makransky et al., 2016) and with statistical methods to allow for an estimation of full CDI scores (Mayor & Mani, 2019) and total vocabulary sizes (Mayor & Plunkett, 2011), tablet-based word recognition tasks may provide a useful measure of receptive vocabulary skills in the second year of life—and potentially serve as a supplemental and convergent measure of parental reports.

ACKNOWLEDGEMENTS

We are very thankful to Ane Theimann for her help with participant recruitment and data collection. We are thankful to all toddlers and their parents who took part in the study.

ORCID

Chang Huan Lo <https://orcid.org/0000-0002-8262-7258>

Audun Rosslund <https://orcid.org/0000-0002-2646-8053>

Jun Ho Chai <https://orcid.org/0000-0003-4316-9407>

Julien Mayor <https://orcid.org/0000-0001-9827-5421>

Natalia Kartushina <https://orcid.org/0000-0003-4650-5832>

REFERENCES

- Abdul Aziz, N. A., Mat Sin, N. S., Batmaz, F., Stone, R., & Chung, P. W. H. (2014). Selection of touch gestures for children's applications: Repeated experiment to increase reliability. *International Journal of Advanced Computer Science and Applications*, 5(4), 97–102. <https://doi.org/10.14569/IJACSA.2014.050415>
- Ackermann, L., Lo, C. H., Mani, N., & Mayor, J. (2020). Word learning from a tablet app: Toddlers perform better in a passive context. *PLoS One*, 15(12), e0240519. <https://doi.org/10.1371/journal.pone.0240519>

- Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, 105(1–2), 63–80. <https://doi.org/10.1016/j.jecp.2009.10.002>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bergelson, E., & Aslin, R. N. (2017a). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, 114(49), 12916. <https://doi.org/10.1073/pnas.1712966114>
- Bergelson, E., & Aslin, R. N. (2017b). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development*, 13(4), 481–501. <https://doi.org/10.1080/15475441.2017.1324308>
- Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, 37(6), 1461–1476. <https://doi.org/10.1017/S0142716416000060>
- Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, 35(3), 619–650. <https://doi.org/10.1017/S0305000908008714>
- Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer* [Computer software]. <http://www.praat.org/>
- Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, 3, 52–67. https://doi.org/10.1162/opmi_a_00026
- Chai, J. H., Lo, C. H., & Mayor, J. (2020). A Bayesian-inspired item response theory-based framework to produce very short versions of MacArthur–Bates communicative development inventories. *Journal of Speech, Language, and Hearing Research*, 63(10), 3488–3500. https://doi.org/10.1044/2020_JSLHR-20-00361
- Clark, R. (1974). Performing without competence. *Journal of Child Language*, 1(1), 1–10. <https://doi.org/10.1017/S0305000900000040>
- Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient delay at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, 46(3), 544–560. [https://doi.org/10.1044/1092-4388\(2003/044\)](https://doi.org/10.1044/1092-4388(2003/044))
- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, 56(8), 848–856. <https://doi.org/10.1111/jcpp.12378>
- Dunn, D. M. (2018). *Peabody picture vocabulary test-fifth edition*. Pearson Assessment.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*, 71(2), 310–322. <https://doi.org/10.1111/1467-8624.00146>
- Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. Singular.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, 21(1), 95–116. <https://doi.org/10.1017/S0142716400001053>
- Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, 83(1), 203–222. <https://doi.org/10.1111/j.1467-8624.2011.01692.x>
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98. <https://doi.org/10.1037/0012-1649.42.1.98>
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231. <https://doi.org/10.1111/1467-9280.00044>
- Frank, M. C., Braginsky, M., Marchman, V. A., & Yurovsky, D. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. MIT Press.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. <https://doi.org/10.1017/S0305000916000209>

- Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development, 17*(1), 1–17. <https://doi.org/10.1080/15248372.2015.1061528>
- Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, & Computers, 35*(2), 302–309. <https://doi.org/10.3758/bf03202556>
- Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): Data from American English and Mexican Spanish infants. *Journal of Child Language, 35*(1), 77–98. <https://doi.org/10.1017/s0305000907008264>
- Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the computerized comprehension task: Comprehension predicts production. *Developmental Psychology, 48*(1), 136–148. <https://doi.org/10.1037/a0025511>
- Friend, M., Smolak, E., Liu, Y., Poulin-Dubois, D., & Zesiger, P. (2018). A cross-language study of decontextualized vocabulary comprehension in toddlerhood and kindergarten readiness. *Developmental Psychology, 54*(7), 1317. <https://doi.org/10.1037/dev0000514>
- Friend, M., & Zesiger, P. (2011). Une réplication systématique des propriétés psychométriques du Computerized Comprehension Task dans trois langues [A systematic replication of the psychometric properties of the CCT in three languages]. *Enfance, 63*(3), 329–344. <https://doi.org/10.4074/S0013754511003053>
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23–45. <https://doi.org/10.1017/s030500090001271x>
- Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science, 8*(3), 316–339. <https://doi.org/10.1177/1745691613484936>
- Hendrickson, K., Mitsven, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. *Developmental Science, 18*(5), 723–735. <https://doi.org/10.1111/desc.12250>
- Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Language, speech, and communication. Methods for assessing children's syntax* (pp. 105–124). MIT Press.
- Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language, 34*(4), 701–724. <https://doi.org/10.1017/s0305000907008124>
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6), F31–F39. <https://doi.org/10.1111/j.1467-7687.2008.00768.x>
- Kaler, S. R., & Kopp, C. B. (1990). Compliance and comprehension in very young toddlers. *Child Development, 61*(6), 1997–2003. <https://doi.org/10.2307/1130853>
- Kartushina, N., & Mayor, J. (2019). Word knowledge in six- to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society Open Science, 6*(9), 180711. <https://doi.org/10.1098/rsos.180711>
- Klesty, V., & Fouche, G. (2020, March 24). Norway extends coronavirus curbs until April 13. *Reuters*, <https://www.reuters.com/article/us-health-coronavirus-norway-restriction/norway-extends-coronavirus-curbs-until-april-13-idUSKBN21B2ED>
- Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means (Version 1.4.5)* [R package]. <https://CRAN.R-project.org/package=emmeans>
- Lo, C. H., Mani, N., Kartushina, N., Mayor, J., & Hermes, J. (2021). e-Babylab: An open-source browser-based tool for unmoderated online developmental studies. *PsyArXiv*. Under review. <https://doi.org/10.31234/osf.io/u73sy>
- Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory-based, computerized adaptive testing version of the MacArthur–Bates Communicative Development Inventory: Words & Sentences (CDI:WS). *Journal of Speech, Language, and Hearing Research, 59*(2), 281–289. <https://doi.org/10.1044/2015.JSLHR-L-15-0202>
- Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., Lahmar, J., Scott, F., Davenport, A., Davis, S., French, K., Piras, M., Robinson, P., Thornhill, S., & Winter, P. (2015). *Exploring play and creativity in pre-schooler's use of apps: Final project report*. <http://www.techandplay.org/reports/TAPFinalReport.pdf>

- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Mayor, J., & Mani, N. (2019). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behavior Research Methods*, 51(5), 2248–2255. <https://doi.org/10.3758/s13428-018-1146-0>
- Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, 14(4), 769–785. <https://doi.org/10.1111/j.1467-7687.2010.01024.x>
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, 86(5), 1351–1370. <https://doi.org/10.1111/cdev.12398>
- Neumann, M. M., Worrall, S., & Neumann, D. L. (2019). Validation of an expressive and receptive tablet assessment of early literacy. *Journal of Research on Technology in Education*, 51(4), 326–341. <https://doi.org/10.1080/15391523.2019.1637800>
- Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, 17(1), 57–70. <https://doi.org/10.1177/1367006911431198>
- Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L. I., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, 24(9), 675–678. <https://doi.org/10.1016/j.tics.2020.06.004>
- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, 34(1), 3–23. <https://doi.org/10.1177/0142723713510997>
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *afex: Analysis of factorial experiments (Version 0.27-2)* [R package]. <https://CRAN.R-project.org/package=afex>
- Styles, S., & Plunkett, K. (2009). What is ‘word understanding’ for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, 36(4), 895–908. <https://doi.org/10.1017/S0305000908009264>
- Syrnyk, C., & Meints, K. (2017). Bye-bye mummy – Word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, 35(2), 202–217. <https://doi.org/10.1111/bjdp.12157>
- Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, 59(5), 174–179. <https://doi.org/10.1111/j.1540-5834.1994.tb00186.x>
- Twomey, D. M., Wrigley, C., Ahearne, C., Murphy, R., De Haan, M., Marlow, N., & Murray, D. M. (2018). Feasibility of using touch screen technology for early cognitive assessment in children. *Archives of Disease in Childhood*, 103(9), 853–858. <https://doi.org/10.1136/archdischild-2017-314010>
- Vales, C., & Fisher, A. V. (2019). When stronger knowledge slows you down: Semantic relatedness predicts children’s co-activation of related items in a visual search paradigm. *Cognitive Science*, 43(6), e12746. <https://doi.org/10.1111/cogs.12746>
- Williams, K. T. (2018). *Expressive vocabulary test-third edition*. Pearson Assessment.
- Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology*, 6(1), 59–64. <https://doi.org/10.1044/1058-0360.0601.59>

How to cite this article: Lo CH, Rosslund A, Chai JH, Mayor J, Kartushina N. Tablet assessment of word comprehension reveals coarse word representations in 18–20-month-old toddlers. *Infancy*. 2021;00:1–21. <https://doi.org/10.1111/inf.12401>

APPENDIX 1

APPENDIX 2