



# Vocabulary Assessment With Tablets in Grade 1: Examining Effects of Individual and Contextual Factors and Psychometric Qualities

Emilie Dujardin<sup>1,2\*</sup>, Jean Ecalle<sup>1</sup>, Pauline Auphan<sup>1,2</sup>, Christophe Gomes<sup>2</sup>, Laurent Cros<sup>2</sup> and Annie Magnan<sup>1</sup>

<sup>1</sup> Laboratory EMC Etude des Mécanismes Cognitifs – MSH LSE USR CNRS 2005 – Université Lyon 2, Lyon, France,

<sup>2</sup> Association Agir pour l'Ecole, Paris, France

## OPEN ACCESS

### Edited by:

Okan Bulut,  
University of Alberta, Canada

### Reviewed by:

Sebastian Paul Suggate,  
University of Regensburg, Germany  
April Lynne Zenisky,  
University of Massachusetts Amherst,  
United States

### \*Correspondence:

Emilie Dujardin  
emilie.dujardin@univ-lyon2.fr

### Specialty section:

This article was submitted to  
Assessment, Testing and Applied  
Measurement,  
a section of the journal  
Frontiers in Education

**Received:** 04 February 2021

**Accepted:** 30 March 2021

**Published:** 27 April 2021

### Citation:

Dujardin E, Ecalle J, Auphan P,  
Gomes C, Cros L and Magnan A  
(2021) Vocabulary Assessment With  
Tablets in Grade 1: Examining Effects  
of Individual and Contextual Factors  
and Psychometric Qualities.  
*Front. Educ.* 6:664131.  
doi: 10.3389/feduc.2021.664131

A tablet application was designed to assess children's receptive vocabulary in French using a classical four-choice picture paradigm and 240 words which varied in word frequency. Results showed (1) an effect of socio-demographic zone, with lower correct response scores and longer reaction times for children in disadvantaged areas, (2) an effect of word frequency, with higher correct response scores and shorter reaction times for frequent words than for rare words, and (3) an effect of age and gender on correct responses in favor of girls and older children. More interestingly, an interaction effect on correct responses revealed that for rare words, the difference between girls and boys was higher, again in favor of girls, in the normal socio-demographic zones. We used an Item Response Theory analysis to examine the psychometric qualities of each item. This then allowed us to select two shortened equivalent versions of the test which were very closely matched to certain psychometric properties. In the same way as other reading-related skills assessed using new technologies (computer or tablet), receptive vocabulary with its two parameters of speed and accuracy can be integrated as an important component of reading ability.

**Keywords:** receptive vocabulary, children, tablet, assessment, item response theory

## INTRODUCTION

The last two decades of research have brought about an increasing interest in increasing the use of mobile phones and touch-screen tablets for assessing, collecting, and developing early literacy skills (see Frank et al., 2016; Herodotou, 2018 for recent reviews). One of these literacy skills is vocabulary. Vocabulary knowledge underpins all language skills, and most specifically reading, and is considered to be an important component of academic success (Rice and Hoffman, 2015). Vocabulary is the knowledge an individual has about a word, and the retrieval of this knowledge is underpinned by processes controlling the speed and access to this knowledge (Oakhill et al., 2012). Vocabulary is often conceptualized in two dimensions: receptive and expressive (see Pearson et al., 2007 for a review). Expressive

vocabulary refers to known words that the individual is able to produce, used correctly in a context/sentence. Receptive vocabulary refers to words that an individual is able to understand, either orally or in writing. It is on the latter that we will focus in this article. In general, the size of receptive vocabulary is larger than the size of expressive vocabulary (i.e., understand more words than they use, Pearson et al., 2007). Applications that test receptive vocabulary knowledge (i.e., understanding a word by reading or hearing it) are easy to develop (Neumann and Neumann, 2019), whereas expressive vocabulary assessment needs implementing more complicated a vocal recording.

The use of digital tools seems to bring advantages in assessment. Indeed, computers or tablets are attractive to children due to their fun aspect (e.g., game), especially for children with learning difficulties or disabilities (see Marble-Flint et al., 2019 for a study in children with autism spectrum disorder). Moreover, these new technologies permit the standardized administration of the tasks and, more importantly, the automated collection of a series of data, response types (correct responses and errors) and response times (Frank et al., 2016; Neumann and Neumann, 2019). It is now generally acknowledged that two parameters, namely speed and accuracy, are important when assessing language skills (Oakhill et al., 2012, 2019; Richter et al., 2013). Children do not all learn at the same speed, do not all have the same vocabulary, and therefore do not all have the same proficiency of language. These inter-individual differences may be due to different factors (e.g., gender, age socio-economic level, or lexical frequency). The aim of this study was a) to use a touch-screen tablet to examine the receptive vocabulary of French-speaking children in Grade 1 as a function of certain individual, contextual and word-related factors, and b) to evaluate the psychometric properties of this test. We briefly present how receptive vocabulary is measured in the literature and then introduce a number of factors affecting vocabulary performance.

## How to Assess Receptive Vocabulary?

Receptive vocabulary is traditionally assessed by means of paper-based tests using a four-choice picture paradigm in which children are asked to match a target word to one of the four pictures. The most widely used and best known test of receptive vocabulary is the *Peabody Picture Vocabulary Test* (in American English, PPVT-4, Dunn and Dunn, 2007), which has been adapted for use in many other languages, such as British English (*British Picture Vocabulary Scale*, Dunn et al., 2009) and French (*Échelle de Vocabulaire en Images Peabody* [*Peabody Picture Vocabulary Scale*], Dunn et al., 1993). Recently, attention has turned to increasing the use of tablets for learning and assessing language skills, especially in children with language difficulties (Marble-Flint et al., 2019). This growing interest in tablet devices has made it possible to develop applications for assessing vocabulary using, for instance, a form of the picture-choice paradigm in multilingual children (Schaefer et al., 2019) as well as in monolingual English-speaking children (Schaefer et al., 2016). The authors pointed out several benefits of using numerical/digital tools to measure vocabulary such as speed, ease of test administration for the experimenter, but ease of

use and comprehension for the children. Another benefit of using digital tools in language assessment is the collection of accurate measures of reaction times and errors. These measures are important because they represent the efficiency of a process (e.g., Richter et al., 2013). Reaction time operationalizes how quickly an individual will access information (see De Boeck and Jeon, 2019 for a review), and errors reflect the accuracy of the information (Oakhill et al., 2012; Richter et al., 2013). In addition, measuring these two indicators also helps determine response strategies, i.e., whether participants react quickly but make errors, or whether they react more slowly and make fewer errors. Using digital tools to assess literacy skills is a new challenge.

## What Factors May Affect Vocabulary Level?

Paper-based assessments of participants' individual vocabulary-related characteristics have been conducted (e.g., Dunn and Dunn, 1981, 1997; Fenson et al., 2007; Taylor et al., 2013; Rice and Hoffman, 2015). A gender effect has been shown, but there is no clear consensus about this. Thirty years ago, Dunn and Dunn (1981) noted that boys performed better than girls both on the original PPVT and on the PPVT-Revised. However, in a more recent study, girls were found to have a larger vocabulary size than boys (Fenson et al., 2007). A longitudinal study, from 2;6 to 21 years of age, showed a gender effect on vocabulary growth with an advantage for young girls which then leveled out with age before turning into an advantage for boys from ages 10 to 21 years (Rice and Hoffman, 2015). The gender effect seems to change as individuals become older. Researchers have not yet provided a clear explanation of this effect (see Rice and Hoffman, 2015, for an explanation in terms of hormones). However, it seems likely that it is partially explained by children's interest in language (e.g., as the interest in reading varies according to gender and/or age; see Hoff, 2006).

Moreover, Rice and Hoffman (2015) showed an effect of age, with the rate of vocabulary acquisition increasing up to 12 years before slowing again. Vocabulary increases with age, and therefore with educational level (Taylor et al., 2013). Vocabulary and reading abilities have a bidirectional relation. Indeed, the more words individuals know, the better their reading comprehension is; and the more they read, the more their vocabulary grows (e.g., Oakhill et al., 2019). One of the main factors explaining vocabulary growth is thus reading practice. Moreover, reading comprehension involves knowing the meaning of words and retrieving this information accurately and quickly when reading. The richness of a person's vocabulary can therefore depend on the speed at which words are activated.

Finally, Taylor et al. (2013) examined receptive vocabulary development, measured with a short version of the PPVT-III (Dunn and Dunn, 1997), in a sample of 4,332 Australian children from 4 to 8 years. They found a negligible effect of gender, but showed that the fact of coming from a disadvantaged socioeconomic area (i.e., families with low socioeconomic status) was related to a lower rate of growth in receptive vocabulary. Indeed, the difference between children from families with high vs. low socioeconomic status derived from the quality of the

conversation between them and the other members of their families (e.g., Hart and Risley, 1995; Fletcher and Reese, 2005). Children from families of a higher socioeconomic status hear more differentiated words (e.g., nouns, verbs, adjectives, etc.) than children from families with lower socioeconomic status (Hoff, 2006). Socioeconomic status affects children's language development in different ways.

Another factor affects vocabulary performance during assessments and it is important to take account of word characteristics, and especially word frequency, when considering vocabulary (Schaefer et al., 2016). Word frequency measures how often a word occurs in an individual's daily life. The more frequent a word is, the better it is known and, therefore, the more likely it is to be part of children's vocabulary. Including rare and frequent words in a test allows varying the difficulty of access to the information of the words stocked in the lexicon. The retrieval of the lexical representation of frequent word is faster and more accurate than that of a rare word. Word knowledge also depends on reading time: The more exposed to reading children are, including shared book reading activities in younger children, the more their vocabulary will increase, and the more they will be able to redefine words they already know (e.g., Cunningham, 2005; Oakhill et al., 2019). In addition, the quality of lexical representation of a word would depend on links between different levels of lexical representation: orthographic, phonological and semantic (Perfetti and Stafura, 2014). The stronger the links between the different levels of representation, the more precise the quality of the lexical representation. The impact of these four factors—gender, age, socio-economic status and word frequency—on vocabulary performance will be examined in the current study.

## PRESENT STUDY

The receptive vocabulary of French-speaking children was assessed using tablets at the beginning of formal reading and writing instruction, i.e., in Grade 1. To respond at the first aim of the paper, we examined the factors which could impact vocabulary performance, such as the socio-demographic zones of schools (related to the socioeconomic status of the families), individual characteristics (gender, age), and word frequency. We operationalized the older vs. younger children by their date of birth, which were included in either semester 1 or semester 2. We expected that children from schools situated in disadvantaged areas, boys and the younger children would achieve lower scores. Moreover, we also expected to observe an effect of word frequency, with the correct response scores decreasing and the response times increasing from frequent words to rare words. In addition, we examined the influence of distractors in the task using two levels of lexical representations: phonological and semantic. This last point will help determine where the children's difficulties lie, whether the accuracy of the information is based more on semantic or phonological information. Finally, in order to examine the psychometric properties of the vocabulary test, an Item Response Theory (IRT) analysis was run to select the best items and then to construct a new version of the vocabulary test.

Thus, based on the most discriminating items, we will propose two shorter versions of the test.

## METHODS

### Participants

A total of 281 first graders ( $M_{age}^1 = 75.9$  months;  $SD = 5.1$ ) took part in this study. They were assessed at the beginning of the school year (October). They were schooled in two zones<sup>2</sup>, one with major "specific educational needs," i.e., part of the so-called "Réseau d'Éducation Prioritaire" [Priority education network] (REP+; 3 schools;  $n = 102$ ; 50 boys/52 girls), and the other outside of the REP (7 schools;  $n = 179$ ; 86b/93g). All necessary consents were obtained from the parents and academic authorities.

### Material and Procedure

A large number of words (240) were presented during two sessions of 30 min each ( $2 \times 120$  words). These consisted of 173 common nouns, 20 adjectives, and 47 verbs. They were selected from a set of available pictures and were divided into five categories ( $48 \times 5$  words) according to their frequency using the UG1 index from the Manulex database (Lété et al., 2004), going from the most frequent words in C1 to the least frequent words in C5: C1 (466.5–67.78), C2 (67.36–32.01), C3 (31.70–10.78), C4 (10.71–1.22), C5 ( $<0.89$ ).

The children performed a traditional task in which they first heard a word and then saw four pictures on the screen (**Figure 1**). Three distractors were presented for each target word: one had a phonological unit (DPho; syllable or rime) in common with the target word, one was from the same semantic field as the target word (DSem), and the third was unrelated to the target word (strange item; DStr).

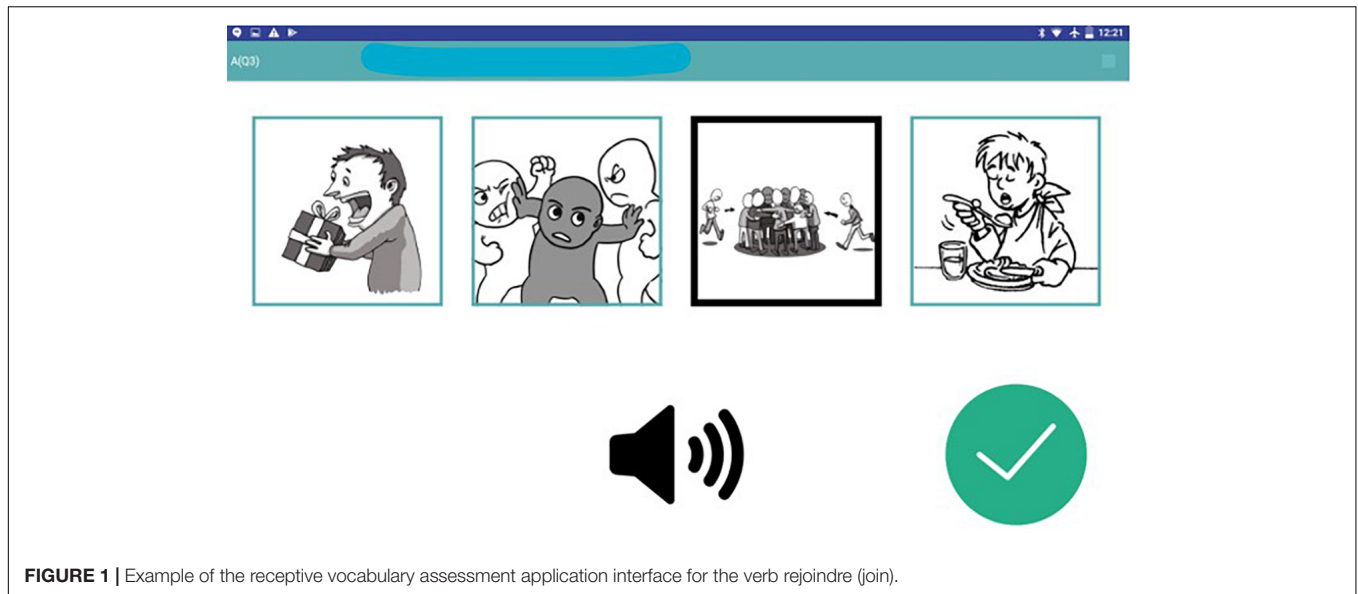
The children sat in front of the tablets wearing headphones. They heard a word and then had to touch the screen with their finger as soon as they wanted to respond (**Figure 1**). If they did not clearly hear the target word, they could ask to hear it again. Of the total responses, 5.94% were collected after a second hearing of the word. The children had 15 s to give each response. After that, a new word was proposed.

The target words were presented randomly and the response time for each response (RT) was recorded. All data (type of response and RT) were saved on a web server and then collected. The test reliability was high with Cronbach's  $\alpha = 0.97$ .

In order to eliminate aberrant response times (outliers), a two-step operation was performed for each participant before the analyses described below were performed: 1/response times that were over two standard deviations from the mean were replaced by the mean response time and 2/after replacement, a mean response time was recalculated.

<sup>1</sup>The age of each child was not recorded.

<sup>2</sup>In the French educational system, schools are divided into three zones according to their socio-economic status and their children's learning difficulties. Two zones with "specific educational needs" are distinguished between: REP and REP+, in which a majority of parents have lower incomes. In REP+, the percentage of manual workers and unemployed persons is the highest (74.3%) compared to that of REP (60%) and non-REP (37.8%) (data published in 2017 by the French Ministry of Education).



**FIGURE 1** | Example of the receptive vocabulary assessment application interface for the verb *rejoindre* (join).

## RESULTS

We first examined the effects of individual, contextual and word-related factors on the collected responses (correct responses and RTs) and types of errors with MANOVA which allows to take into account the effects of the different independent variables on the combination of dependent variables. Then, we conducted IRT analyses to examine the psychometric properties of all the items and select the items sufficiently discriminant to assess receptive vocabulary. After selecting the best items, that is to say those with good properties, we tried to construct two shortened versions of the test which exhibited the same properties to avoid a long test administration time.

### Effects of Individual, Contextual and Textual Factors

Descriptive data are presented in **Table 1**. Two successive MANOVAs with the same design were run, one on the correct response scores and the second on the RT, with three between-subjects factors, Zone (REP+ vs. outside of REP), Gender (boys vs. girls) and Age (old, born in semester 1 vs. young, born in semester 2) and one within-subjects factor, Word Frequency (C1, C2, C3, C4, and C5).

With regard to the correct response scores, we observed significant effects of Zone,  $F(1, 266) = 24.95$ ,  $p < 0.0001$ ,  $\eta^2 = 0.09$ , with children outside of REP outperforming those from REP+, of Age,  $F(1, 266) = 10.97$ ,  $p = 0.001$ ,  $\eta^2 = 0.04$ , with the older children achieving higher correct response scores than the younger ones, and of Word Frequency,  $F(1, 266) = 201.16$ ,  $p < 0.0001$ ,  $\eta^2 = 0.43$ , with the mean correct response scores decreasing from C1 to C5, although we did observe a plateau effect between C4 and C5. A significant interaction between Zone, Gender, and Word Frequency was found,  $F(4, 1064) = 2.89$ ,  $p = 0.02$ ,  $\eta^2 = 0.02$  (**Figure 2**). For the least frequent words (C4

and C5), the differences between girls and boys increased as a function of Zone.

For the dependent variable, namely RTs, a significant effect of Zone,  $F(1, 266) = 52.86$ ,  $p < 0.0001$ ,  $\eta^2 = 0.17$  was found. The RTs were higher for children outside of REP than for children from REP+. Moreover, we again found an expected effect of Word Frequency,  $F(4, 1064) = 76.40$ ,  $p < 0.0001$ ,  $\eta^2 = 0.22$ , with RTs decreasing from C1 (frequent words) to C5 (rare words).

Finally, we carried out a MANOVA on the types of errors, with a between-subjects factor Zone and a within-subjects factor Type of Error, for the three distractors: phonological (DPho), semantic (DSem) and strange (DStr). We expected a significant interaction between Zone and Type of Error, with the differences between the children in the two zones being greater for DSem and DPho. A significant effect of Zone was again revealed,  $F(1, 279) = 33.04$ ,  $p < 0.0001$ ,  $\eta^2 = 0.11$ , with the children from REP+ making more errors than their peers outside of REP (31.4 vs. 23.3). We also found a significant effect of Type of Error,  $F(2, 588) = 510.98$ ,  $p < 0.0001$ ,  $\eta^2 = 0.65$ , with the number of errors decreasing from DPho (32.6) to DSem (32.1) to DStr (17.4). There was no significant difference between the first two of these and the expected significant interaction was not found ( $p > 0.05$ ). No other significant effects were found.

### Psychometric Properties: An IRT Analysis

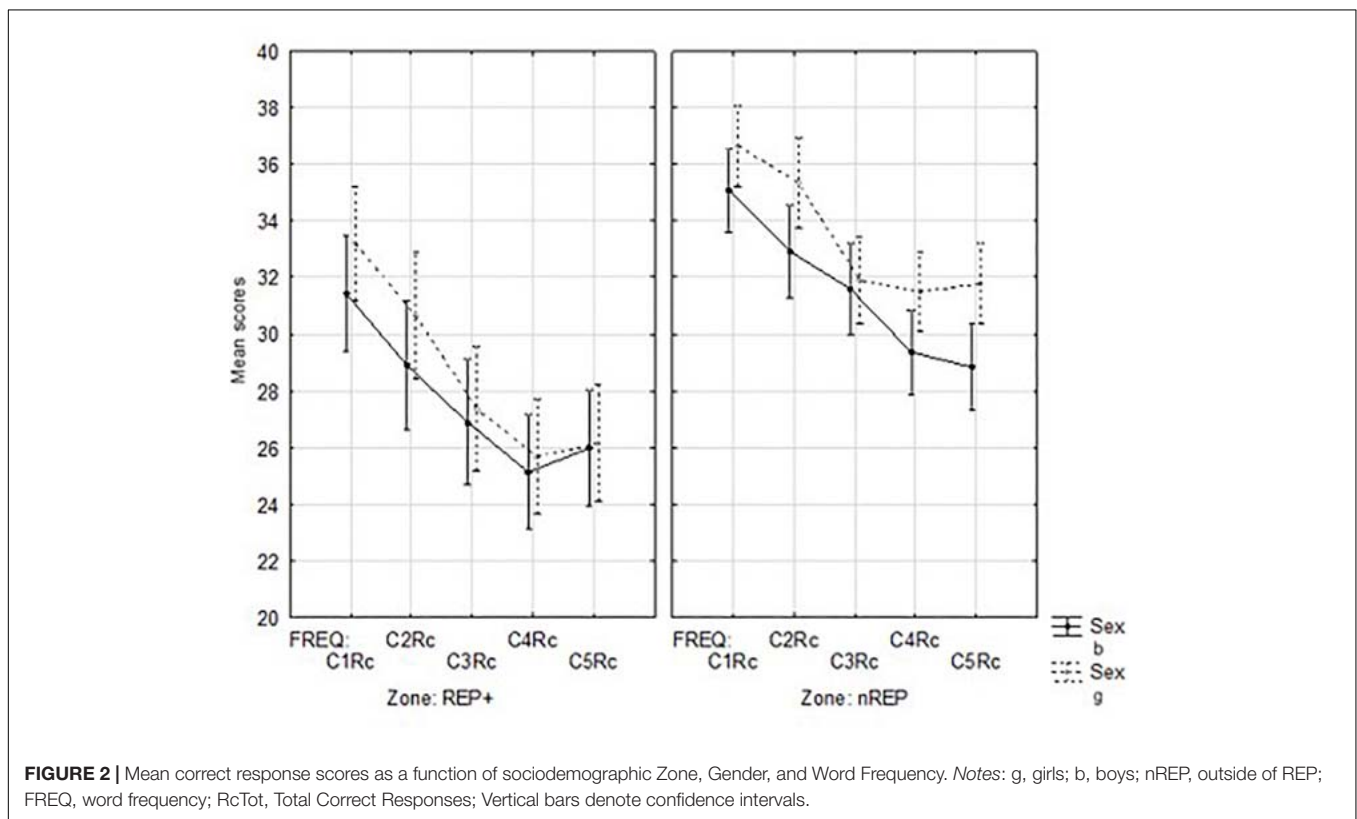
We conducted an IRT analysis using a two-parameter logistic model (2PL) to obtain the difficulty and discrimination coefficients of the items. Here, we present two types of curves, namely item characteristics curves (icc) in **Figure 3A**, and the test characteristic curve (tcc) in **Figure 3B**. In the first graph (iic), high-discrimination items are represented by a steep slope and items with a flat slope are poorly discriminated. A negative discrimination coefficient (see “*luire*,” **Figure 3A**) indicates that the corresponding item was not informative for the purposes



**TABLE 1 |** Mean (standard deviations in brackets; Skewness and Kurtosis) correct response scores and response times as a function of zone, gender, age and word frequency.

Correct response scores		/240					/48				
Zone		Gender		Age		Frequency					
REP+	nREP	boys	girls	sem1	sem2	C1	C2	C3	C4	C5	
142.2	162.6	150.8	159.3	161.6	149.1	34.7	32.7	30.2	28.8	28.9	
(32.4)	(35.5)	(37.3)	(33.8)	(33.2)	(37.3)	(7.2)	(8)	(7.9)	(7.4)	(7.5)	
-0.25	-0.75	-0.42	-0.5	-0.6	-0.34	-0.80	-0.63	-0.4	-0.42	-0.19	
2.42	2.9	2.34	2.58	2.78	2.28	1.91	2.86	2.5	2.63	2.27	
Response times											
2,925	3,806	3,478	3,494	3,433	3,586	3,300	3,371	3,518	3,608	3,633	
(889)	(905)	(1,025)	(972)	(929)	(1,051)	(960)	(996)	(1,034)	(1,055)	(1,059)	
1.1	-0.07	0.21	0.23	-0.11	0.36	0.21	0.42	0.29	0.30	0.21	
4.04	2.91	2.41	2.49	2.01	2.55	2.36	2.94	2.68	2.71	2.51	

nREP, outside of REP; sem1, children with birthdays in semester 1; sem2, children with birthdays in semester 2.

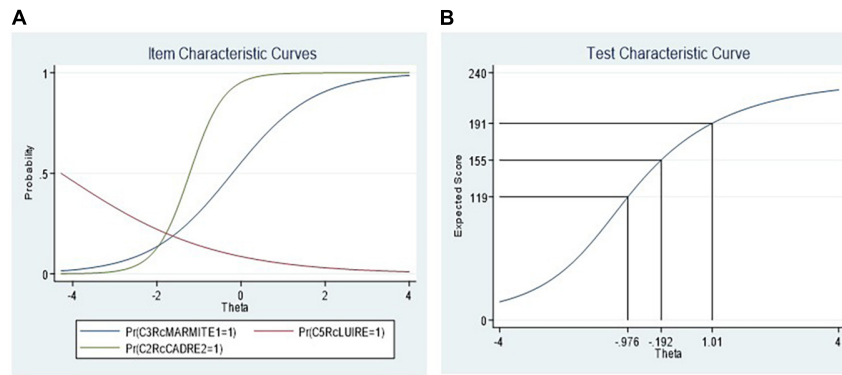


of the test (Baker, 2001). The second graph (tcc) covers all the items. It is obtained by adding, for each theta value ( $\theta$ ), the probabilities relative to all the items. **Figure 3B** presents three contrasted  $\theta$  values corresponding to three expected scores (correct responses), the mean, the mean less one standard deviation and the mean plus one standard deviation. The more accentuated curve on the left side (negative  $\theta$ ) shows that the test was more difficult for the children with a lower vocabulary level (expressed as the latent trait).

### Selection of Items for Two New Versions of the Test

A two-step process was used: 1/we calculated the point-biserial coefficient ( $r_{pb}$ ) of each item and 2/we constructed two versions with items matched on their difficulty coefficients from the IRT analysis.

Taking a  $r_{pb} < 0.25$  as our threshold, 42 items were discarded. These items were also those with the lowest discrimination coefficients (and obviously those with negative values). We then



**FIGURE 3 |** (A) icc for three contrasted items as a function of their discrimination coefficients, indicating the lowest ( $\alpha = -0.55$ ; *luire*; shine), the highest ( $\alpha = 2.45$ ; *cadre*; frame) and a mean coefficient discrimination ( $\alpha = 1.03$ ; *marmite*; pot); (B) tcc with three expected scores (mean score ( $m = 155$ ),  $m\text{-}sd = 119$ ), and  $m + sd = 191$ ) and the corresponding latent trait scores ( $N = 281$ ). Notes: icc, item characteristic curves; tcc, test characteristic curve; m, mean; sd, standard deviation.

**TABLE 2 |** Mean correct response scores (with standard deviations and range in brackets) per item, difficulty and discrimination coefficients as a function of versions of test.

	Version A	Version B
Item correct responses	0.700 (0.16) 0.30 to 0.94	0.695 (0.16) 0.31 to 0.94
Difficulty coeff.	-0.954 (0.82) -2.54 to 1.31	-0.957 (0.82) -2.63 to 0.31
Discrimination coeff.	1.205 (0.39) 0.57 to 2.05	1.205 (0.44) 0.58 to 2.40

*Coeff.*, coefficient.

matched two series of items on the basis of their difficulty coefficients. We finally obtained two versions ( $2 \times 99$  items) of the test with very similar difference indexes (Table 2). Moreover, after conducting an IRT analysis for each version, we can observe that the two curves (Figure 4) have very similar slopes and that the theta values are also very similar for the average correct response scores in the two versions. Finally, with regard to the two tcc, we can confirm that versions A and B have the same properties.

## DISCUSSION

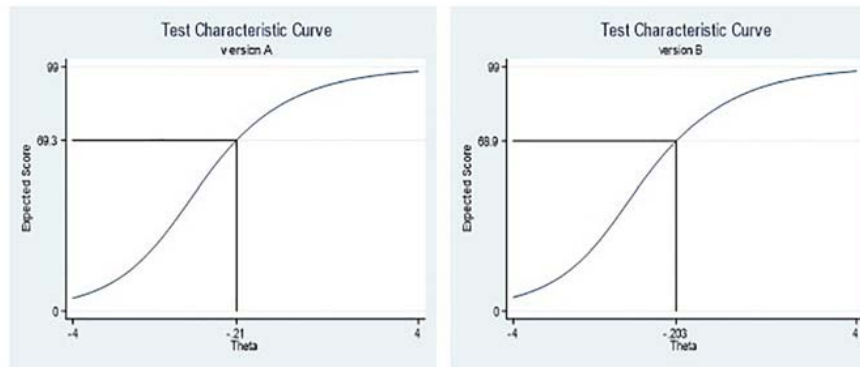
The aim of this paper was to use tablets to assess the receptive vocabulary level of French-speaking children at the beginning of Grade 1 as a function of individual, contextual and word-related factors and to examine the psychometric properties of this new test implemented on a tablet. We used a conventional, easy four-picture choice task.

We found an effect of educational zone, with children outside of REP performing better than those from REP+. The children in REP+ came from families with low socioeconomic status. They made more errors than the children from outside of REP and exhibited shorter RTs. This latter finding might seem surprising. Indeed, we might have expected the RTs of the children from

lower socio-economic zones to be longer because their less developed vocabulary knowledge should have caused them to hesitate when responding. In the light of these unexpected faster RTs combined with more errors, we assume that the children in REP+ may ultimately have exhibited speed over accuracy. At the same time, the poorer receptive vocabulary abilities of the lower socioeconomic status children (i.e., REP+ in our study) is an expected result (see, for example, Taylor et al., 2013). According to Hoff (2006), this could result from the different interaction between these children and their parents. Families with low socioeconomic status might use a different language style (e.g., less complex grammar and syntax in families with lower socioeconomic status). Moreover, children from families with higher socioeconomic status read (with their parents) more often than those from families with lower socioeconomic status (e.g., Fletcher and Reese, 2005). As a result, children from families with higher socioeconomic status are more likely to understand and know words and develop language abilities, and more specifically improve their vocabulary knowledge (e.g., Hoff, 2006; Oakhill et al., 2019).

We also found an effect of age. The older children had lower RTs and higher correct response scores than their younger counterparts in the same school year. These results confirm that receptive vocabulary develops rapidly in early childhood (e.g., Taylor et al., 2013; Rice and Hoffman, 2015), especially at the beginning of formal educational. Indeed, we show an age effect when only a few months separate the birth of children (i.e., born in the first half of the year vs. born in the second half of the year).

Furthermore, receptive vocabulary performances vary as a function of word frequency. RTs decreased and the number of correct responses increased with increasing word frequency. In this study, we used a frequency index calculated in G1 (Lété et al., 2004). The retrieval of the lexical representation of a frequent word is faster and more accurate than that of a rare word. When children with a low level of vocabulary read low-frequency words then, even if they can decode the words they read, one of the dimensions of the words will be missing, i.e.,



**FIGURE 4 |** Test characteristic curves for  $m = 69.25$  in version A and for  $m = 68.87$  in version B.

the semantic representation (see Perfetti and Stafura, 2014), and this will hinder or prevent reading comprehension. Vocabulary should therefore be assessed as a core component of reading ability and word frequency is an important factor that needs to be taken into account when developing a corresponding assessment tool.

More interestingly, an interaction effect between gender, zone and word frequency was found. The word frequency effect was larger for girls than for boys as a function of socioeconomic zone (i.e., REP+ vs. outside of REP), with girls outperforming boys on rare words in normal zones (outside of REP). Our results are consistent with those of Fenson et al. (2007) and Rice and Hoffman (2015), showing that girls have better language abilities than boys (see Hoff, 2006, for a review). In our study, we did not have information about the reading practices of the children with their parents. That is to say, we were unaware of their “home literacy environment” (see Sénéchal et al., 2017). It might be worthwhile including this factor in future work because research shows that children with a rich home literacy environment (books, magazines, shared book-reading, etc.) are more likely to develop vocabulary (Hart and Risley, 1995). Our study confirms that receptive vocabulary knowledge depends on different factors, which may be individual (e.g., gender, age), contextual (e.g., socioeconomic zone), or related to word characteristics (e.g., word frequency).

The design of our test allowed us to observe how the types of error vary depending on the relation shared with the target item (e.g., semantic, phonological, or strange—not related to the target). We thus examined the types of errors made by the children and this enabled us to identify and understand their difficulties. The children made more “phonological” and “semantic” errors than “strange” errors. The “phonological” errors may be accounted for by an auditory attentional bias. The children might have made errors because they did not concentrate enough to respond accurately, or they might not have heard the word clearly. However, as we have pointed out, they were able to listen to each word again. The “semantic” errors are due more obviously to imprecise knowledge of the target word (see Perfetti and Stafura, 2014). Finally, the “strange” errors that

we observed would seem to suggest that the target words were completely unknown.

Another important result of our study is that we can now propose two versions of a French receptive vocabulary test available in the form of a tablet application. Indeed, the initial test might have made use of too many items and was therefore time-consuming. We are now in a position to shorten this version to produce two truncated versions which could be used as complementary tools in further studies. For instance, one version could be used at the beginning of an interventional study, and the second at the end of the study in order to avoid or attenuate the classical test-retest effect. These two versions have the same good psychometric properties. The IRT analyses have shown that the discrimination coefficients are sufficiently high to identify children with low or high receptive vocabulary abilities.

Finally, using a tablet has various advantages, which have been repeatedly highlighted in the literature (Frank et al., 2016; Schaefer et al., 2016, 2019; Herodotou, 2018; Neumann and Neumann, 2019). Indeed, digital tools are generally popular with children and increase their motivation. Moreover, the receptive format test is easy to administer and easy to understand for children. It could also be used with atypical populations of children, e.g., those with an autism spectrum disorder (Marble-Flint et al., 2019), with specific language impairments, or with difficulties in learning to read in the case of children at the beginning of Grade 1. In addition, the use of computerized tools in language assessment allows for the collection of accurate response times and errors, which are two essential indicators to consider when estimating the efficiency of language processes (i.e., the speed of access to and accuracy of the information requested—e.g., Oakhill et al., 2012; Richter et al., 2013).

To conclude, our article provides a digital vocabulary assessment tool that is fun and easy to administer to children. We have provided two equivalent versions with several important items, which will allow the measurement of receptive vocabulary without any test-rest effect. We found effects of individual (i.e., gender, age), contextual (i.e., socioeconomic zone), and lexical (i.e., word frequency) variables. These variables should be considered in the development of future norms.

## LIMITATIONS AND FURTHER RESEARCH

First, not all children have the same prior experience with tablets. Children who have already used a tablet are more likely to be confident in indicating their responses to the tests. It may therefore be necessary to take some time to familiarize children with the tablet (Frank et al., 2016). Second, we had no information about home language, the home literacy environment or maternal education. However, these factors could influence the level of receptive language (Schaefer et al., 2016; Sénéchal et al., 2017; see Hoff, 2006 for a review). Third, given that vocabulary has two dimensions, i.e., breadth and depth (Ouellette, 2006), the second of these will be developed in the next version of our test. A more complete version, including the two dimensions, would then be available for administration to older children in subsequent grades. Using the same linguistic material as is employed in computer-based assessments of reading, word reading (Auphan et al., 2019), and comprehension (Beauvais et al., 2018), vocabulary will be another important reading-related skill which can be examined in order to define the profiles and difficulties of readers. In such an approach, it will be possible to take account of both parameters, namely speed and accuracy. Of course, to obtain a good and reliable all-round tool for assessing both reading and vocabulary, all the tasks will need to be implemented on tablets. Furthermore, comparison with a paper/pencil vocabulary test should be tested to confirm the advantage of the computerized version (Neumann and Neumann, 2019). In addition, future work will propose a standardized norm of this vocabulary test associated with a reading test implemented on tablets. Another limitation is that the sample is composed solely of Grade 1 children.

## REFERENCES

- Auphan, P., Ecalle, J., and Magnan, A. (2019). Computer-based assessment of reading ability and subtypes of readers with reading comprehension difficulties: a study in French children from G2 to G9. *Eur. J. Psychol. Educ.* 34, 641–663. doi: 10.1007/s10212-018-0396-7
- Baker, F. (2001). *The Basics of Item Response Theory. Clearinghouse on Assessment and Evaluation*. College Park, MD: University of Maryland.
- Beauvais, L., Bouchafa, H., Beauvais, C., Kleinsz, N., Magnan, A., and Ecalle, J. (2018) Tinfolec: a new French web-based test for reading assessment in primary school. *Can. J. Sch. Psychol.* 33, 227–241. doi: 10.1177/0829573518771130
- Cunningham, A. E. (2005). “Vocabulary growth through independent reading and reading aloud to children,” in *Teaching and Learning Vocabulary: Bringing Research to Practice*, eds E. H. Hierbert and M. L. Kamil (Lawrence: Erlbaum).
- De Boeck, P., and Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Front. Psychol.* 10:102.
- Dunn, A., and Dunn, A. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., and Dunn, D. M. (1997). *Peabody Picture Vocabulary Test 3rd Edition*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., and Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. London: Pearson Assessments.
- Dunn, L. M., Dunn, D. M., and Styles, B. (2009). *The British Picture Vocabulary Scale*. London: GL Assessment Limited.
- Dunn, L. M., Dunn, D. M., and Theriault-Whalen, C. (1993). *EVIP: Échelle de Vocabulaire en Images Peabody [Pictures Vocabulary Scale Peabody]*. Canada: Pearson Assessment.

Standards applicable to a larger population (i.e., at different levels and/or by age) should be the subject of future studies. This will then provide a standard of reference for researchers, practitioners, and teachers.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

JE, AM, LC, CG, and PA contributed to the design and development of the study. LC and CG provided access to the study population. JE organized the database and performed the statistical analysis. PA, AM, JE, and ED participated in the interpretation of the results, as well as in the choice of theory. ED wrote the introduction and the discussion parts of the manuscript. JE wrote the method and results parts of the manuscript. All authors contributed to the revision of the manuscript, read and approved the submitted version.

- Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories, User's Guide and Technical Manual*. Baltimore, MD: Brookes.
- Fletcher, K. L., and Reese, E. (2005). Picture book reading with young children: a conceptual framework. *Dev. Rev.* 25, 64–103. doi: 10.1016/j.dr.2004.08.009
- Frank, M. C., Sugarman, E., Horowitz, E. C., Lewis, M. L., and Yurovsky, D. (2016). Using tablets to collect data from young children. *J. Cogn. Dev.* 17, 1–17. doi: 10.1080/15248372.2015.1061528
- Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: P.H. Brookes.
- Herodotou, C. (2018). Young children and tablets: a systematic review of effects on learning and development. *J. Comput. Assist. Learn.* 34, 1–9. doi: 10.1111/jcal.12220
- Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002
- Lété, B., Sprenger-Charolles, L., and Colé, P. (2004). MANULEX: a grade-level lexical database from French elementary school readers. *Behav. Res. Methods Instrum. Comput.* 36, 156–166. doi: 10.3758/bf03195560
- Marble-Flint, K. J., Strattman, K. H., and Schommer-Aikins, M. A. (2019). Comparing iPad® and paper assessments for children with ASD: an initial study. *Commun. Disord. Quart.* 40, 152–155. doi: 10.1177/1525740118780750
- Neumann, M. M., and Neumann, D. L. (2019). Validation of a touch screen tablet assessment of early literacy skills and a comparison with a traditional paper-based assessment. *Int. J. Res. Method Educ.* 42, 385–398. doi: 10.1080/1743727x.2018.1498078
- Oakhill, J., Cain, K., and Elbro, C. (2019). “Reading comprehension and reading comprehension difficulties,” in *Reading Development and Difficulties: Bridging*



- the Gap Between Research and Practice*, eds D. A. Kilpatrick, R. M. Joshi, and R. K. Wagner (Cham: Springer).
- Oakhill, J., Cain, K., McCarthy, D., and Field, Z. (2012). "Making the link between vocabulary knowledge and comprehension skill," in *From Words to Reading for Understanding (101–114)*, eds A. Britt, S. Goldman, and J.-F. Rouet (Milton Park: Routledge).
- Ouellette, G. P. (2006). What's meaning got to do with it: the role of vocabulary in word reading and reading comprehension. *J. Educ. Psychol.* 98, 554–566. doi: 10.1037/0022-0663.98.3.554
- Pearson, P. D., Hiebert, E. H., and Kamil, M. L. (2007). Vocabulary assessment: what we know and what we need to learn. *Read. Res. Quart.* 42, 282–296. doi: 10.1598/rrq.42.2.4
- Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Sci. Stud. Read.* 18, 22–37. doi: 10.1080/10888438.2013.827687
- Rice, M. L., and Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: a longitudinal study from 2;6 to 21 years of age. *J. Speech Lang. Hear. Res.* 58, 345–359. doi: 10.1044/2015\_jslhr-14-0150
- Richter, T., Isberner, M.-B., Naumann, J., and Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Sci. Stud. Read.* 17, 415–434. doi: 10.1080/10888438.2013.764879
- Schaefer, B., Bowyer-Crane, C., Herrmann, F., and Silke Fricke, S. (2016). Development of a tablet application for the screening of receptive vocabulary skills in multilingual children: a pilot study. *Child Lang. Teach. Ther.* 32, 179–191. doi: 10.1177/0265659015591634
- Schaefer, B., Ehlert, H., Kemp, L., Hoesl, K., Schrader, V., Warnecke, C., et al. (2019). Stern, Gwiazda or star: screening receptive vocabulary skills across languages in monolingual and bilingual german–polish or german–turkish children using a tablet application. *Child Lang. Teach. Ther.* 35, 25–38. doi: 10.1177/0265659018810334
- Sénéchal, M., Whissell, J., and Bildfell, A. (2017). "Starting from home: home literacy practices that make a difference," in *Studies in Written Language and Literacy*, eds K. Cain, D. L. Compton, and R. K. Parrila (Amsterdam: John Benjamins Publishing Company).
- Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., and Zubrick, S. R. (2013). Risk factors for children's receptive vocabulary development from four to eight years in the longitudinal study of australian children edited par K. Paterson. *PLoS One* 8:e73046. doi: 10.1371/journal.pone.0073046

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Dujardin, Ecalle, Auphan, Gomes, Cros and Magnan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.