Infant Behavior and Development 57 (2019) 101379

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

Full length article

Do early lexical skills predict language outcome at 3 years? A longitudinal study of French-speaking children

Tamara Patrucco-Nanchen[a,*], Margaret Friend[b], Diane Poulin-Dubois[c], Pascal Zesiger[a]

[a] FPSE, University of Geneva, Switzerland
[b] San Diego State University, USA
[c] Concordia University, Canada

ARTICLE INFO

Keywords:
Language acquisition
Lexical comprehension
Prediction
Toddlers

ABSTRACT

Early language development is considered critical for children's adjustment in school, for social adaptation and for later educational achievement. Despite the role of children's receptive skills as a foundation for later productive word use, receptive language skills have received surprisingly little attention. The present research extends recent work on the prediction of preschool language skills by exploring whether a decontextualized measure of lexical comprehension can account for unique variance in preschool language skills above and beyond parent report and how early such a prediction can be made. For this purpose, 65 French-speaking children have been tested at 16, 22, 29 and 36 months.

The results of the current study suggest that up to the age of two, although parent reports of lexical comprehension and/or production account for a portion of variance in later receptive, productive or general language outcome, they have less predictive validity than a direct measure of early lexical comprehension. By contrast, after age two, parent reported vocabulary production is the strongest predictor of later language production skills.

1. Introduction

Oral language abilities depend on a variety of skills including phonological, semantic, pragmatic, social, and cognitive that interact with one another (Bates et al., 1994; Hoff, 2006). These skills are considered critical for children's adjustment in school (Law, Rush, Parsons, & Schoon, 2013; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015), for social adaptation and development (Botting & Conti-Ramsden, 2008; Fujiki, Brinton, Isaacson, & Summers, 2001; Snowling, Bishop, Stothard, Chipchase, & Kaplan, 2006), and for later educational achievement (Snowling, Adams, Bishop, & Stothard, 2001). It is therefore important to identify reliable measures that can serve as predictors of later language skills so that possible difficulties can be detected or prevented. The primary aim of the present paper is to compare the relative efficacy of a direct measure of early receptive vocabulary (the Computerized Comprehension Task; CCT) and a parent report checklist, the *Inventaire Français du Développement Communicatif* (IFDC), the French adaptation of the *MacArthur-Bates Communicative Development Inventory*, (MCDI), for predicting language skills at the group level in French-speaking children. The second aim is to evaluate the efficacy of these measures as a function of the timing of measurement (i.e., 16 vs. 22 months and 22 vs. 29 months) for the prediction of specific dimensions of language outcome at 36 months.

* Corresponding author at: Tamara Patrucco-Nanchen, Université de Genève, Boulevard du Pont d'Arve 28, 1211 Genève 4, Switzerland.
E-mail address: tamara.patrucco@unige.ch (T. Patrucco-Nanchen).

https://doi.org/10.1016/j.infbeh.2019.101379
Received 13 November 2018; Received in revised form 12 July 2019; Accepted 8 September 2019
Available online 24 September 2019
0163-6383/

1.1. Parent reports

Early language skills are usually assessed using parent reports (like the MCDI, Fenson et al., 1993) that document lexical comprehension and production. The MCDI is a measure of early language and communicative development that is cost effective, easy to administer, adapted into many languages, and is used in cross-linguistic investigations, as well as in clinical and research settings. It is efficient and valid in young children both with and without developmental disabilities. It is also a rich source of information on early communicative and language acquisition (for a review, see Law & Roy, 2008).

On the one hand, various studies have shown that parents are competent at evaluating their child's production of words regardless of maternal education, income, and ethnic background (e.g. Feldman et al., 2005). Recent research using parent reports have shown that by 2 years of age, the size of expressive vocabulary predicts later language abilities (Duff, Reen, Plunkett, & Nation, 2015; Ghassabian et al., 2014; Henrichs et al., 2011; Kemp et al., 2017; Lee, 2010; Marchman & Fernald, 2008a; Reilly et al., 2010), school readiness (Friend, Smolak, Liu, Poulin-Dubois, & Zesiger, 2018), reading skills (Duff, Reen et al., 2015; Lee, 2010), academic and behavioral functioning (Morgan et al., 2015), social-emotional adjustment (Irwin, Carter, & Briggs-Gowan, 2002), and cognitive abilities (Marchman & Fernald, 2008a). However, parent reports appear to have limited predictive power for language delays (Duff, Nation, Plunkett, & Bishop, 2015; Ghassabian et al., 2014), and explain only a small proportion of outcome variance (Duff, Reen et al., 2015; Henrichs et al., 2011; Lee, 2010). Consequently, their usefulness as a predictor that has clinical significance has been questioned (Duff, Reen et al., 2015, Duff, Nation et al., 2015; Henrichs et al., 2011; Morgan et al., 2015; Reilly et al., 2010; Rescorla, 2011).

On the other hand, early lexical comprehension skills have been predicted to be more predictive of later language skills, such that poor receptive language skills are associated with disadvantaged socioeconomic circumstances during childhood, and with more behaviour and psychosocial adjustment problems in the transition to adulthood (Schoon, Parsons, Rush, & Law, 2010). Several recent studies have shown a link between poor language comprehension and language delay (e.g. Henrichs et al., 2011; Ellis Weismer, 2007; Zambrana, Pons, Eadie, & Ystrom, 2014). More precisely, these studies indicate that poor language comprehension at 18 months of age, assessed by parent report, predicts expressive vocabulary delay at the group level, and is the most important of early predictors. Finally, several studies show that most children with both a receptive and an expressive lexical delay at 2 years of age display oral language and academic difficulties throughout childhood and adolescence, and are at risk of developing learning disorders as well as cognitive, behavioural and psychiatric difficulties (Ellis Weismer, 2007; Henrichs et al., 2011; Law, Boyle, Harris, Harkness, & Nye, 2000; Rescorla, 2009; Rice, Taylor, & Zubrick, 2008; Snowling et al., 2001, 2006).

However, it has also been argued that parent reports of vocabulary comprehension are less reliable than those of vocabulary production (Houston-Price, Mather, & Sakkalou, 2007; Styles & Plunkett, 2008), not sufficiently consistent over time (Yoder, Warren, & Biggar, 1997), and not sufficiently predictive of developmental outcomes at the individual level (Feldman et al., 2005; Zambrana et al., 2014).

1.2. Direct language measures

There is a need to develop reliable measures of infants' receptive skills that can serve as predictors of later language abilities. However, given the challenge of assessing language directly and objectively in very young children, only a few studies have actually addressed this issue. To our knowledge, there are only three direct measures designed to assess infants' receptive skills before the age of 24 months: the Intermodal Preferential Looking paradigm (IPL; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987); the Looking-While-Listening paradigm (LWL; Fernald, Perfors, & Marchman, 2006; Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998), and the Computerized Comprehension Task (CCT; Friend & Keplinger, 2003, 2008). The IPL focuses on infants' language comprehension, the LWL focuses on word processing efficiency, whereas the CCT focuses on receptive vocabulary size.

The IPL uses the cumulated looking time to a target and a distractor image to assess lexical comprehension (Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996). The IPL paradigm has been used to investigate several key questions about infants' emerging language including phonology, lexicon, grammar and morphology (see Golinkoff, Ma, Song, & Hirsh-Pasek, 2013 for a review), but not as a tool to predict infants' later language skills.

The LWL uses real-time measures of the time course of toddlers' gaze patterns in response to speech, and assesses infants' word processing efficiency. In early studies Fernald et al. (Fernald et al., 2001; Zangl, Klarman, Thal, Fernald, & Bates, 2005) provided preliminary evidence for a relation between speech processing and vocabulary size during the second year of life. In a longitudinal study, Fernald et al. (2006) reported a strong link between processing efficiency and vocabulary size by showing that children who responded more quickly and reliably in the LWL task at 25 months of age showed greater acceleration in vocabulary growth across the second year. By following a subsample of the initial cohort up to the age of 8 years, Marchman and Fernald (2008a) have also shown that children's speech processing efficiency and productive vocabulary size at 25 months strongly predicted performance on measures of school-age language and cognitive abilities. Finally, speech processing, which accounted for 4%–17% additional variance in vocabulary growth trajectories of 30-month-old late-talkers, improves the ability to predict persistent delays over and above expressive vocabulary size alone (Fernald & Marchman, 2012). Although the LWL procedure "is low in task demands and does not require automated eye-tracking technology, similar to "preferential looking" procedure…," it requires "meticulous procedures in the collection, reduction, and multiple levels of analysis of such details data are demanding" (Fernald, Zangl, Portillo, & Marchman, 2008, p.97). Thus, although the procedure is highly informative with regard to early receptive language processing, this labour-intensive approach reduces its portability for the purposes of assessing children outside the lab.

With these concerns in mind, Friend and Keplinger (2003, 2008) developed a direct measure of infants' lexical comprehension, the Computerized Comprehension Task (CCT) that is now available in English, French, and Spanish. The CCT operationalizes vocabulary size as the number of discrete haptic responses to a referent, and is a reliable and valid measure of comprehension vocabulary in the second year (Friend & Keplinger, 2008). This direct measure of decontextualized receptive vocabulary (Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015) can be used from 16 to 24 months of age and has been shown to demonstrate comparable psychometric properties across 3 languages (Friend & Zesiger, 2011). The CCT has also the advantage of being a portable, easy to administer assessment and to be appropriate for the attention limitations of children in the second year of life (Friend & Keplinger, 2008). Recent studies using the CCT have shown that, before 2 years of age, receptive vocabulary size uniquely accounts for 24% of variance (13% when controlling for parent-reported vocabulary) in vocabulary diversity in a language sample in the third year of life (Friend, Schmitt, & Simpson, 2012), for 20 to 25% of variance in general preschool language outcome at 3 years of age (Friend, Smolak, Patrucco-Nanchen, Poulin-Dubois, & Zesiger, 2018), and for 24% of variance in children's kindergarten readiness at 4 years of age (Friend, Smolak, Liu et al., 2018).

1.3. Alternative language measures

Even if it is widely assumed that early screening for a range of developmental outcomes is necessary and efficient, Duff, Reen et al. (2015) suggest that the measurement of language skills "at 3 years of age is still too early to be reliably informative at an individual level" (p. 853). In contrast, Henrichs et al. (2011) conclude that screening for early speech and language problems should be carried out after the age of 18 months. Other authors (Conti-Ramsden & Durkin, 2012; Fenson et al., 1993; Henrichs et al., 2011) encourage the use of supplemental measures assessing both receptive and expressive skills in the various dimensions of the language system to explain additional variance in language outcome.

Language assessment of pre-schoolers usually cover vocabulary and grammar in comprehension and production (Conti-Ramsden & Durkin, 2012). In addition to standardized measures assessing those skills, two tools (Redmond, Thompson & Goldstein, 2011) have recently received particular attention in the language acquisition literature: non-word repetition and sentence repetition. Non-word repetition (for a review, see Coady & Evans, 2008) was designed to assess phonological short-term memory (Conti-Ramsden & Durkin, 2012). In this task, the participant is asked to repeat non-words containing an increasing number of syllables. Accuracy in repeating non-words has previously been found to correlate with language outcome both in typically developing children and in children with developmental language disorders (Adams & Gathercole, 2000; Bishop, North, & Donlan, 1996; Botting & Conti-Ramsden, 2001). More recently, sentence repetition has been demonstrated to be a window into the participants' language competence (Marinis & Armon-Lotem, 2015) and morphosyntactic skills (Polišenská, Chiat, & Roy, 2015), and argued to be more sensitive and specific than non-word repetition in identifying monolingual children with language impairment (Botting & Conti-Ramsden, 2001; Stokes, Wong et al., 2006).

Finally, language acquisition in pre-schoolers can also be assessed in a more ecological way by using language sample analysis. In their review, Heilmann, Miller, and Nockerts (2010) point out that language sample analysis has been used in research on language development for over 50 years, and has been the basis for most of our current knowledge on typically developing children's language production. Samples of conversational language offer an accurate picture of a child's production skills in a natural context both at the lexical and grammatical levels (Trudeau, 2007). In summary, the literature on language acquisition in pre-school children indicates that, in order to obtain a comprehensive profile of the language skills at different ages, it is useful to assess short-term phonological memory capacities as well as receptive and expressive lexical and grammatical skills using different measures (direct, indirect) and, if possible, in different contexts.

2. The current study

The first goal of our study is to test whether direct and indirect measures of vocabulary acquisition collected at two different time points during the second year of life each contribute for unique variance in language skills assessed at two years and a half. We therefore assessed typically developing French-speaking monolingual children on the MCDI and on the CCT at ages 16 and 22 months and used these scores to predict their language outcome at 29 months, which has also been assessed by a combination of direct and indirect measures. Following the results reported by Friend et al. (2012); 2018b), we expect that the CCT, being a direct and decontextualized measure of early lexical comprehension, will account for a larger part of variance of the language outcome measure than parental reports that are typically used at this age.

The second aim of our study is to quantify the additional variance in language outcome at 3 years explained by a set of measures assessing both receptive and expressive language skills at 29 months. For this purpose, we included both spontaneous and elicited measures of language assessing lexical and grammatical comprehension, lexical and grammatical production, and non-word repetition accuracy. At 29 months, we anticipate that non-word repetition accuracy (see Coady & Evans, 2008 for a review) will be the major predictor of language production outcome. The study protocol was approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences of the University of Geneva.

2.1. Study 1a

This first part of the study assesses the predictive power of direct (CCT) and indirect (MCDI) measures collected at two points in time (16 and 22 months) on language outcome at 29 months.

2.1.1. Material and methods

2.1.1.1. Participants. Sixty-six monolingual French-speaking children (33 girls and 33 boys) were recruited through birth lists provided by the State of Geneva in Switzerland. Data were collected longitudinally and participants visited the laboratory for three visits. Out of the 66 children, one was excluded due to a technical problem, two to attrition, and 13 because they were unable to complete the testing in one of the three first sessions. The final sample consisted of 50 toddlers (23 girls and 27 boys) at 16 months of age (χ = 15.99, σ = .33; range 15.2–17.02) at Wave 1, 22 months (χ = 21.94, σ = .28; range 21.01–22.21) at Wave 2 and 29 months (χ = 28.71, σ = .42; range 28.13–29.19) at Wave 3. The mean duration of the mothers' education was 15.26 years (σ = *2.73*; range 11–22). At each visit, the experimenter filled out the Language Exposure Assessment Tool (LEAT; DeAnda, Bosch, Poulin-Dubois, Zesiger, & Friend, 2016), a computer-based assessment that estimates both quantitative and qualitative aspects of language exposure in the child's daily life on the basis of an intensive parent interview. Mean exposure to French was.97 (σ = .05; range .82–1) at Wave 1,.96 (σ = .05 range .80–1) at Wave 2, and.95 (σ = .06; range .82–1) at Wave 3.

2.1.1.2. Measures at Waves 1 and 2 (16 and 22 months)

2.1.1.2.1. Inventaire Français du Développement Communicatif (IFDC). The IFDC (Kern, 1999) is the European-French adaptation of the MacArthur-Bates Communicative Development Inventory (MCDI, Fenson et al., 1993). The *Mots et Gestes* (MG) form is designed for children from eight to sixteen months of age. It includes a list of 414 words organised in 19 semantic categories. Parents indicate whether their child understands (IFDC-comprehension score) and/or produces (IFDC-production score) these words. The *Mots et Phrases* (MP) form is designed for children from sixteen to thirty months. It consists of a list of 690 words organised in 22 semantic categories. Parents indicate whether their child produces these words. They are also asked to report examples of the longest utterances (IFDC-MLU score) produced by their child. The MG form was used at Wave 1, and the MP form, at Wave 2.

2.1.1.2.2. Computerized Comprehension Task (CCT). The French adaptation (Friend & Zesiger, 2011) of the CCT (Friend & Keplinger, 2003; 2008) is a software program designed to directly assess lexical comprehension in toddlers. It is administered as a forced-choice procedure on a touch screen. The child is asked to touch the correct picture out of two pictures shown side by side on the screen following the prompts (e.g. Where is the dog? Touch *dog*! Who is sleeping? Touch *sleeping*! Which one is blue? Touch *blue*!). Each trial lasts a maximum of 7 s. If the child touches the target image, a reinforcing auditory signal is delivered (repetition of the word and target-related sound).

The CCT contains vibrantly-colored, high-quality digital images which are matched within trials for brightness, size, color, word class and word difficulty. Word difficulty was based on normative data from the IFDC: MG at 16 months of age (Kern & Gayraud, 2010), with equal numbers of easy (comprehended by more than 66%), moderately difficult (comprehended by 33%–66%) and difficult (comprehended by less than 33%) words. The task includes 4 training trials, 41 test trials (23 pairs of nouns, 11 pairs of verbs and 7 pairs of adjectives), and 13 reliability trials. In each trial, pairs of pictures which belong to the same semantic category are presented simultaneously on left- and right-center of a touch screen. There are two forms of the CCT such that each word serves as both a target and a distractor. Forms were counterbalanced across participants.

During the task, the child was seated on the parent's lap approximately 30 cm from the screen where they could touch it easily. Parents were asked to wear sound-attenuating headphones and darkened glasses (or close their eyes) to prevent interference during the administration of the CCT. The examiner sat to the right of the parent and presented pairs of items by touching an invisible button on the bottom right corner of the touch screen. She administered 4 training trials with easy words and repeated these training trials until the child demonstrated that they understood the nature of the task by correctly responding to at least one of the training trials. Next the examiner began the administration of the test trials.

The CCT score corresponds to the number of correct responses on the task (out of 41 items). Responses were coded from a video offline. Reliability checks were performed for 36% (N = 18) of the 16-month-olds and 26% (N = 13) of the 22-month-olds and yielded inter-agreements of .98 (range = .88–1) for Wave 1 and.96 (range = .93–1) for Wave 2.

2.1.1.3. Measures at Wave 3 (29 months)

2.1.1.3.1. MP form of the IFDC. (Kern, 1999) identical to Wave 2. IFDC-production score was calculated.

2.1.1.3.2. The lexical and grammatical comprehension subtests of the Evaluation du Langage Oral (ELO, Khomsi, 2001). These subtests are designed for children from 2 ½ to 10 years of age. The lexical comprehension task is composed of 20 test items, and that of the grammatical comprehension task, of 2 training items followed by 20 test items. The measure yields two outcome variables: the child is asked to point to the picture out of four choices that best matches the meaning of a target word (ELO-words score) or utterance (ELO-sentences score).

The order of administration of these 2 tasks was counterbalanced across participants.

2.1.1.4. Procedure. In order to allow for later coding, all tasks were recorded in full with two Panasonic WV-CS320-G Colour Dome cameras and two AKG CK62 ULS microphones connected to a Data Video Mixer SE 800. Children were tested individually in a quiet room in our laboratory in the presence of a parent. At the beginning of each visit, the parent was informed about the tasks to be performed with the child, and asked to fill out a consent form and a short demographic questionnaire covering the child's health history, changes in the language input, and possible concerns about the child's language development. Once the parent had finished completing the questionnaires, the test session started and lasted approximately 45 min for Wave 1 and Wave 2, and 60 min for Wave 3.

At the end of each visit, the parents were asked to complete the IFDC at home the same day and return it to the lab by mail. They received a $25 gift card in a bookshop and infants received a small gift, either a toy or book, for their participation.

2.1.2. Results

All variables met the criteria of a normal distribution, with the exception of the IFDC produced words, at 16 months. At this age, the IFDC comprehension score ranged from 52 to 387 words (*M = 204.9, SD = 76.31),* and the IFDC production score ranged from 1 to 185 words (*M = 30.04, SD = 35.33)*. The skewness (2.48) and kurtosis (7.312) coefficients of this latter variable indicated that it did not meet the criterion of a normal distribution reflecting the fact that most children produced few words. Due to the asymmetry observed in this distribution, this variable was transformed using a natural logarithm function in the subsequent analyses. The receptive vocabulary score on the CCT ranged from 2 to 32 words (*M = 16.08*, *SD = 7.06*). Internal consistency across forms was excellent (Cronbach's α = .927 and .954 for forms A and B, respectively). Eighteen children completed the reliability phase of the assessment and stability in performance across test and reliability phases was moderate ($r(18) = .55$, $p = .019$).

At 22 months, the IFDC production score ranged from 13 to 523 words (*M = 220.74, SD = 142.18)* and the IFDC Maximum Length of Utterances (IFCD-MLU) ranged from 1 to 10.67 words (*M = 3.38, SD = 1.95)*. The receptive vocabulary score on the CCT ranged from 12 to 40 words (*M = 29.4*, *SD = 6.37*). The internal consistency of the CCT was excellent with Cronbach's α = .934 and .937 for forms A and B, respectively. Forty-three children completed the reliability phase of the assessment and stability in performance across test and reliability phases was strong ($r(43) = .63$, $p < .001$).

At 29 months, the IFDC production score ranged from 40 to 646 words (*M = 446.4, SD = 155.35)* and the IFDC-MLU ranged from 2.33 to 18.67 words (*M = 8.31, SD = 4.14)*. The ELO receptive vocabulary score (ELO-words) ranged from 5 to 15 (*M = 10.9, SD = 2.26*) and the ELO grammar comprehension score (ELO-sentences) ranged 4 to 16 from (*M = 11.42, SD = 2.9*).

2.1.2.1. Relations between early lexical measures and language abilities at 29 months. Preliminary correlational analyses showed that there was no correlation between either the child's gender *(all ps > .159)* or mother's education (*all ps > .064*) and any of the language measures for the three waves. Zero-order correlations between predictors and outcome variables were computed. As can be seen in Table 1, all measures of lexical development at 22 months correlate with two or more language outcome measures at 29 months. However, the CCT score is the only 22 months variable that correlates with all four 29 months language outcome measures.

2.1.2.2. Predictors of language abilities. In order to eliminate method variance (Bornstein, Hahn, & Putnick, 2016), we computed a factorial analysis with the *Z*-score of all language measures at age 29 months, that is to say the IFDC-production score, the IFDC-MLU, the ELO-words score and the ELO-sentences score. The KMO value (.751) and Bartlett's test ($X^2(6) = 52.32$, $p < .001$) indicated that the factor analysis was useful for these variables. A single factor was extracted representing 59.72% of the variance (see Table 2 for component matrix).

This Language Factor significantly correlated with the IFDC-comprehension score ($r(50) = .282$, $p = .047$) at 16 months, and with the IFDC-production score ($r(50) = .542$, $p < .001$), IFDC-MLU score ($r(50) = .446$, $p < .001$) and the CCT score ($r(50) = .709$, $p < .001$) at 22 months, but neither with the child's gender *(p = .463)* nor with the mother's education (p = *.581*). These latter two variables were consequently not included in the subsequent analyses.

2.1.2.2.1. Which measures of lexical development, including direct (CCT) and parent report (IFDC) assessments at two time points (16 and 22 months), best predict language abilities at 29 months in typically developing children?. Since this research question aims at identify the best predictor(s) among measures that have been assessed at two different time points, we chose to use hierarchical multiple linear regressions to predict the 29 months Language Factor as a dependent variable. The predictors were entered in two steps: the *Z*-scores of each language variable assessed at 16 months (IFDC-production, IFDC- comprehension, and CCT score) were entered in Step 1, and those measured at 22 months (IFDC-production, IFDC-MLU and CCT score) were entered in Step 2. Within each step, in order to identify the strongest predictor(s), the stepwise method was used (Howell, 1997). Tolerance values were above .595 at every step indicating low collinearity between the variables.

Results indicate that the IFDC-comprehension score at 16 months, $F\Delta(1,48) = 5.137$, $p = .028$, $R^2\Delta = .079$, and both the CCT

Table 1
Bivariate correlations for all predictors and outcome variables for Study 1a.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Maternal Education											
2. Sex	.030										
3. IFDC-comprehension score 16 m.o.	−.264	.012									
4. IFDC-production score 16 m.o	−.132	.047	.429**								
5. CCT score 16 m.o	−.093	−.097	.237	.079							
6. IFDC-production score 22 m.o.	−.041	.065	.222	.352*	.187						
7. IFDC-MLU score 22 m.o.	−.038	.083	.212	.306*	.176	584**					
8. CCT score 22 m.o.	−.107	.202	.141	−.113	.418**	.257	.375**				
9. IFDC-production score 29 m.o	.074	.089	.245	.119	.246	.649**	.500**	.440**			
10. IFDC-MLU score 29 m.o	.051	.096	.223	.031	.220	.399**	.425**	.607**	.647**		
11. ELO-words score 29 m.o	.080	.041	.221	.073	.186	.358*	.269	.523**	.435**	.509**	
12. ELO-sentences score 29 m.o.	.043	.103	.179	−.043	.194	.235	.207	.644**	.410**	.426**	.318*

Note: IFDC, Inventaire Français du Développement Communicatif; CCT, Computerised Comprehension Task; ELO, Evaluation du Langage Oral, subtests of receptive skills.

* p < .05.

** p < .01.

Table 2
Component Matrix of the Language factor extracted after the factorial analysis with language measures at 29 months for Study 1a.

Measure	Component 1
IFDC-production score 29 m.o.	.824
IFDC-MLU score 29 m.o.	.856
ELO-words score 29 m.o.	.725
ELO-sentences score 29 m.o.	.672

Extraction Method: Principal Component Analysis.
Note: 1 component extracted; IFDC, Inventaire Français du Développement Communicatif; CCT, Computerised Comprehension Task; ELO, Evaluation du Langage Oral, subtests of receptive skills.

score, $F\Delta(1,47) = 42.413$, $p < .001$, $R^2\Delta = .432$, and the IFDC-production score, $F\Delta(1,46) = 16.344$, $p < .001$, $R^2\Delta = .125$, at 22 months explained a total of 65.6% of the variance of the Language factor at 29 months, $F(3,49) = 28.597$, $p < .001$. See Table 3 for all regression parameters.

2.1.2.2.2. Which model provides the best fit to the data?. Finally, we contrasted the three models that emerged from the hierarchical regressions to select the model with least information loss using the Akaike Information Criterion approach based on Information Theory (Posada & Buckley, 2004). The purpose of this approach is to choose the model with the highest quality taking into account both the goodness of fit and the complexity of the model. In this approach, the model with the lower values is associated with higher quality. The approach yielded AICs of 143.26, 111.26 and 130.42 for Models 1 to 3, respectively. Given our sample size, the observed difference in AIC scores allows us to conclude that the Model 2 (directly assessed receptive vocabulary on the CCT) provides the best fit to the data (Hilbe, 2011).

2.1.3. Discussion

The goal of the first part of the study was to determine which measure of lexical development, including direct (CCT) and parent report (IFDC) assessments at two time points (16 and 22 months) has the greatest potential to predict language abilities at 29 months in typically developing children. Overall, 53.1% of the variance of the final language score was explained by the regression model that provides the best fit to the data, and that includes only the CCT assessed at age 22 months as a significant predictor. In fact, most likely given the common variance that this measure shares with the IFDC comprehension score at 16 months, it suppresses the effect of this earlier predictor, which now only appears as a trend in Model 2.

In Model 3, expressive vocabulary reported by the parents explains 12.5% of the variance of the children's language skills, confirming the predictive power of this variable on later language abilities (Duff, Reen et al., 2015; Kemp et al., 2017). However, the fact that the best fit between time points and measurements is obtained for the model in which the CCT at 22 months was the predictor of interest, together with the fact that this task predicts 43.2% of the total variance of the Language Factor at 29 months, allow us to conclude that the direct comprehension task is the most powerful predictor of later language abilities. This result thus confirms the predictive potential of the CCT on short-term outcomes in different languages (Friend, Smolak, Patrucco-Nanchen et al., 2018, 2012).

Zambrana et al. (2014) reported that poor language comprehension evaluated with the CDI at 17 months was the earliest communication predictor for children with language delay between 3 and 5 years. Other research on parent reports of vocabulary comprehension have shown that this measure is less reliable than vocabulary production (Houston-Price et al., 2007), inconsistent over time (Yoder et al., 1997), and not sufficiently predictive of developmental outcomes (Feldman et al., 2000). In line with these findings, our results show that this indirect measure of lexical comprehension at 16 months is the earliest predictor of later language abilities. However, on the one hand, this contribution is limited (less than 8% of explained variance), and on the other, this early contribution is overridden by a later direct measure of lexical comprehension like the CCT. Note that the CCT, which was also introduced as a predictor at the age of 16 months, did not contribute to explain the language outcome variance at 29 months, which

Table 3
Hierarchical regression parameters for models predicting Language Factor at 29 months from IFDC word's comprehension score at 16 months and CCT score and IFDC word's production score at 22 months for Study 1a.

	Model 1						Model 2						Model 3					
Measure	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*
	.099					**.028**	**.531**					**.000**	**.656**					**.000**
IFDC-word's comprehension W1		**.079**	5.137	.140	.314	**.028**				.104	.180	.087				.093	.092	.320
CCT-score W2								**.432**	42.413	.100	.671	**.000**				.088	.605	**.000**
IFDC-word's production W2														**.125**	16.344	.090	.373	**.000**

Note: IFDC, Inventaire Français du Développement Communicatif; CCT, Computerised Comprehension Task.

indicate that at such an early age, this direct measure may be too complex for most children to perform given their cognitive and attentional limitations. Therefore, parental reports at age 16 months, despite the limitations mentioned above (Houston-Price et al., 2007; Yoder et al., 1997; Feldman et al., 2000), may be more informative than direct measures. Henrichs et al. (2011) reported on the outcome of nearly 4000 Dutch infants 30 months of age whose expressive vocabulary size had been evaluated by parents at 18 months. Their multifactorial model with 15 predictors could only account for 17.7% of variance in expressive vocabulary at 30 months. More specifically, expressive vocabulary scores at 18 months accounted for 11.5% of the explained variance. Friend et al. (2012) showed that parent reports of word comprehension between the age of 16–21 months (M = 18.2 months) accounted for approximately 25% of the variance in parent reported expressive vocabulary at the age of 24–41 months (M = 28.27). However, parent reported comprehension did not predict unique variance in expressive vocabulary when controlling for CCT scores. Our results, in which age ranges at each time point were more constrained, confirm Zambrana et al.'s (2014) study and extend the results of Friend et al.'s (2012) and Henrichs et al.'s (2011) studies. Although, as in Henrichs et al.'s (2011) study, we found that parent reported expressive vocabulary at 22 months accounted for 12.5% of language ability at 29 months, when we incorporated a direct measure of early vocabulary, we were able to explain a much larger proportion of the variance of language outcome (53.1%) using only 2 predictors, and 65.6% using only 3 predictors.

Friend, Smolak, Patrucco-Nanchen et al. (2018) reported that directly assessed lexical comprehension at 22 months in English-speaking and French-speaking children was the most reliable predictor of a composite language score computed on the basis of the results obtained in a word comprehension task, a sentence repetition task, and the MLU of a language sample at age 36 months. Whether this task can predict more specific dimensions of language outcome at 3 years, such as lexical comprehension, and lexical and grammatical production, remains to be tested, which is the purpose of Study 1b.

2.2. Study 1b

The second part of the study aims at testing the predictive power of vocabulary measures at 22 months of age relative to lexical and grammatical skills and phonological short-term memory at 29 months of age to predict receptive and productive language abilities at 36 months.

2.2.1. Material and methods

2.2.1.1. Participants. The same sample in Study 1a, was tested once more at the age of 36 months. One additional child was excluded because she was unable to complete the testing during session 4. The final sample consisted of 49 toddlers (24 girls and 25 boys). The mean age of the toddlers was 21.9 months (σ = .28; range 21.01–22.21) at Time 2, 28.7 months (σ = .42; range 28.13–29.19) at Time 3, and 35.8 months (σ = .48; range 34.28–37.05) at Time 4. Mean French exposure at each time was respectively .97 (σ = .06; range .80–1), .96 (σ = .06; range .82–1) and .96 (σ = .06; range .82–1).

2.2.1.2. Measures at Time 2 (22 months). identical to Study 1a.

2.2.1.3. Measures at Time 3 (29 months)

2.2.1.3.1. MP form of the IFDC. Kern (1999) identical to Study 1a.

2.2.1.3.2. The lexical and grammatical comprehension subtests of the Evaluation du Langage Oral. (ELO, Khomsi, 2001) identical to Study 1a.

2.2.1.3.3. Non-word Repetition Task (NRT). The NRT (procedure based on Hoff, Core, & Bridges, 2008) is considered a powerful tool for testing language performance in typical and impaired children (for a review, see Coady & Evans, 2008). The task was modified in our lab to make it appropriate for French-speaking participants and included 2 training trials using monosyllabic non-words and 12 test trials using 1-, 2- and 3-syllable non-words (four trials per length). The non-words were created using words of the MG form of the IFDC which were transformed following two basic principles: 1) for monosyllabic non-words: monosyllabic words were selected from the IFCD and the first phoneme was changed, 2) for the multisyllabic non-words; they were built by combining syllables existing in words of the IFDC in the same positions (see Table A1 in Appendix A).

The administration of the NRT was based on the procedure developed and validated by Hoff et al. (2008) and Parra, Hoff, and Core (2011). The trials were presented orally by a native French-speaking examiner sitting in front of the child. The non-words were accompanied by toys representing people and animals. During the test, the examiner showed a toy, said the non-word as if it was the toy's name and asked the child to repeat it back (for example: "This guy is named Bam. Can you say *Bam*?"). If the child did not repeat the name, the examiner repeated the non-words up to three times (except for the two training items, in which case the non-word was repeated as many times as necessary).

Only the first repetition produced by the child was scored, regardless of its accuracy. If a child failed to repeat the non-words for 6 consecutive trials, the test was ended. Only children who attempted to repeat at least 3 non-words were included in the analyses. The accuracy of non-word repetition was measured by calculating the total number of phonemes presented that were repeated correctly by the child (NRT score). All children successfully completed the task. The scoring of repetition accuracy was done by French native speakers. Thirty-three percent (N = 16) of non-words repetition data were reliability-coded by a second native coder. Phoneme-by-phoneme analysis yielded inter-rater agreement of .97 and inter-rater agreement on the number of correct non-words was .95.

2.2.1.3.4. Free play language sample. The task consists of the recording of one parent and their child playing with an extended Fisher Price farm play set including the farm, 3 vehicles, 4 people, 10 animals and 7 miscellaneous items. The parent was instructed to play with the child just as they would do at home.

The language samples were transcribed using the Systematic Analysis of Language Transcripts software (SALT, research version, Miller, Iglesias, & Nockerts, 2012) and coded for grammatical morphology followed the Québec-French adaption of SALT conventions developed by Thordardottir (2005). To ensure stability across transcripts, the number of different words (LS-words score) and Mean Length of Utterance in morphemes (LS-MLU score) analysis were restricted to the first 100 complete and intelligible utterances. Reliability checks were performed by a second coder on 25% of the recordings (N = 12). Point-by-point inter-rater agreement for the child language samples was .89 for morpheme-by-morpheme agreement, and .95 for word-by-word agreement.

2.2.1.4. Measures at Time 4 (36 months)

2.2.1.4.1. Free play language sample. Same as Wave 3[1] .

2.2.1.4.2. Échelle de Vocabulaire en Images Peabody (EVIP). The EVIP (Dunn, Dunn, & Theriault-Whalen, 1993) is the normed Canadian French version of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). It is a direct measure of receptive vocabulary for ages 30 months to adulthood and was normed on a large representative sample of French speakers in Canada. The child is instructed to point to one picture out of four that best matches a word given by the examiner. The test contains 4 training trials. The number of test trials varies as a function of the child's performance. The test ends when the child commits 6 errors out of a set of 8 consecutive trials. The EVIP score is the number of items to reach the ceiling minus the number of errors.

2.2.1.4.3. Nouvelles Epreuves d'Evaluation du Langage (NEEL). The NEEL (Chevrie-Muller & Plaza, 2001) is a French standardized battery of language measures. We used only the reduced form of the picture naming subtest, which contains 14 test trials of picture naming for ages 3,7–8,7 years. A score of 2 is awarded if the expected word is produced in the correct phonological form; a score of 1 is awarded if the expected word is produced but contains a mild or moderate phonological alteration; a score of 0 is awarded in all other cases. The scoring of word production accuracy (NEEL score) was done by a French native speaker. Twenty-two percent of word production data (N = 11) were reliability-coded by a second native coder and yielded an inter-agreement score of .90.

2.2.1.4.4. Sentence Repetition Task (SRT). The SRT included 3 simple sentences as training and 27 sentences of different lengths, syntactic complexity, and morphology based on the lexicon of the MG form of the IFDC and following the sentence construction proposed by Devescovi & Caselli, 2007 (see Table B1 in Appendix B).

A picture representing its global meaning accompanied each sentence. During the test, the experimenter said the sentence first, then showed the picture to the child and told him to repeat the sentence. If the child did not repeat, the examiner repeated the sentence a second time without covering the picture (except for the training items, in which case the sentence was repeated as many times as necessary).

Regardless of its accuracy, only the first repetition produced by the child was scored. If a child failed to repeat a sentence on 6 consecutive trials, the test was ended. The accuracy of sentence repetition was measured by calculating the number of morphemes that were correctly repeated (SRT score). The scoring of repetition accuracy was done by French native speakers. All of the sentence repetition data were reliability-coded by a second native coder. Morpheme-by-morpheme analysis yielded inter-rater agreement of .98 and inter-rater agreement on the number of sentences correct was .99.

2.2.1.5. Procedure. To allow for later coding, the same equipment as in Study 1a was used to record the tasks. The same procedure described in Study 1a was used for 22 months-olds. At 29 months, once the parent had finished completing the questionnaires, the test session started and lasted approximatively 45 min. The child and their parent were guided to a carpet on which were placed the Fischer Price farm. They were instructed to play together with it as they would do at home for 20 min. After this playtime, the examiner offered the child and the parent to move to a small table to administer the two subtests of the ELO and the NRT to the child. The order of the 3 tests was counterbalanced over participants. At age 36 months, after the completion of the questionnaires, the test session started and lasted approximately 60 min. The parent and the child were first asked to play with the Fischer Price farm during 15 min. After this playtime, the child and their parent were guided to a small table to administer the EVIP, the subtest of the NEEL battery and the SRT. The order of these 3 tests was counterbalanced over participants. As in Study 1a, at the end of each visit, parents received an equivalent of $25 gift card in a bookshop and infants received a small gift for their participation.

2.2.2. Results

All of the data met the criteria of a normal distribution for the intended analyses.

At 22 months, the IFDC-production score ranged from 13 to 523 ($M = 211.24$, $SD = 142.21$). The CCT score ranged from 12 to 40 words ($M = 29.4$, $SD = 6.37$). The internal consistency of the CCT at Wave 2 was excellent (Cronbach's α = .919 and .915 for forms A and B, respectively). Forty-three children completed reliability trials and test-retest stability was moderate ($r(43) = .526$, $p < .001$).

At 29 months, the IFDC-production score ranged from 40 to 646 words ($M = 440.82$, $SD = 156.10$). The ELO-words score ranged from 5 to 15 ($M = 10.8$, $SD = 2.18$) and the ELO-sentences score ranged 4 to 16 from ($M = 11.39$, $SD = 2.986$). The NRT score (at the phoneme level) ranged from 2 to 56 ($M = 42.80$, $SD = 10.77$). The MLU in morphemes (LS-MLU score) and the number of different words (LS-words score) produced by children during the Free Play ranged from 1.50 to 4.81 ($M = 3.00$, $SD = .77$), and from 27 to 159 ($M = 89.59$, $SD = 23.31$), respectively.

[1] A total of 46 (96%) children at W3 and 48 (98%) children at W4 met the criterion of 100 utterances leaving 3 children whose transcripts included less than 100 utterances (M = 70, range = 69 to 72) at W3 and 1 child whose transcript included only 89 utterances at W4. A case-by-case review indicated that there was no systematic difference in outcomes that distinguished these participants from the larger sample. That is why we retained these cases in the analyses and calculated the number of different words produced (LS-words score) and the Mean Length of Utterance in morphemes (LS-MLU score) over the entire sample.

Table 4
Bivariate correlations for all predictor and outcome variables for Study 1b.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Maternal Education														
2. Sex	−.090													
3. IFDC- production score 22 m.o.	−.042	.033												
4. CCT score 22 m.o.	−.076	.161	.259											
5. IFDC- production score 29 m.o.	.044	.106	.638**	.424**										
6. ELO-words score 29 m.o.	.172	.017	.322*	.490**	.382**									
7. ELO-sentences score 29 m.o.	−.013	.111	.170	.594**	.385**	.231								
8. NRT score 29 m.o.	−.078	−.062	.555**	.457**	.612**	.112	.424**							
9. LS-word score 29 m.o.	.153	.104	.363*	.231	.502**	.246	.050	.390**						
10. LS-MLU score 29 m.o.	.066	.129	.369**	.301*	.492**	.268	.071	.354*	.840**					
11. EVIP score 36 m.o.	−.003	.189	.349*	.525**	.349*	.526**	.444**	.137	.104	.243				
12. NEEL score 36 m.o.	−.019	−.022	.311*	.482**	.609**	.374**	.285*	.571**	.554**	.547**	.190			
13. SRT score 36 m.o.	−.080	−.012	.252	.266	.540**	.246	.320*	.605**	.396**	.401**	.201	.599**		
14. LS-words score 36 m.o.	.102	.194	.329*	.346*	.559**	.413**	.099	.386**	.474**	.500**	.174	.596**	.346*	
15. LS-MLU score 36 m.o.	.050	−.107	.345*	.273	.514**	.314*	.147	.480**	.544**	.580**	.159	.486**	.424**	.755**

Note: IFDC, Inventaire Français du Développement Communicatif; CCT, Computerised Comprehension Task; ELO, Evaluation du Langage Oral, subtests of receptive skills; NRT, Nonword Repetion Task; LS, Language Sample; EVIP, Échelle de Vocabulaire en Images Peabody; NEEL, Nouvelles Epreuves d'Evaluation du Langage, subtest of productive vocabulary ; SRT, Sentence Repetition Task.

* $p < .05$.

** $p < .01$.

At 36 months, the EVIP score and the NEEL score ranged from 7 to 61 (*M = 26.77, SD = 10.76),* and from 2 to 26 (*M = 16.86, SD = 4.83)*, respectively. The SRT score ranged from 25 to 168 (*M = 133.31, SD = 36.58*). The LS-MLU score and the LS-words score ranged from 2.18 to 5.67 (*M = 3.97, SD = .86),* and from 64 to 151 (*M = 11.94, SD = 21.79)*, respectively. In order to standardize the scales of the different indices used, all the scores obtained in the three waves were transformed into *Z*-scores in the subsequent analyses.

2.2.2.1. Relations between early lexical measures and language outcome. As in Study 1a, preliminary analyses showed no significant correlation between the child's gender *(all ps > .219)* or mother's education (*all ps > .182*) and any of the language measures for the three waves. As can be seen in Table 4, zero-order correlation between predictor and outcome variables reveal that all variables collected at 22 and 29 months correlated with three or more language outcome measures. At age 36 months, the EVIP score, a receptive vocabulary measure, is the only variable that does not correlate with any of the other variables.

2.2.2.2. Predictors of language outcomes. In order to eliminate method variance (Bornstein et al., 2016), we computed an exploratory factor analysis to provide a broad estimate of language production skills taking into consideration spontaneous language (which is subject to contextual variation) and standardized assessment tools (which may fail to capture the richness of the child's language) in the lexical and grammatical domains at 3 years of age. We therefore entered the NEEL score (lexical production), the SRT score (grammatical production and more general skills) and the LS-words (lexical diversity production) and the LS-MLU (grammatical production) of the Language Sample. All participants contributed data for each indicator. The KMO value (.650) and Bartlett's test ($X^2(6) = 79.434$, $p < .001$) indicated that the factor analysis was useful for these variables. A single Productive Language Factor explained 64.85% of the variance (see Table 5 for component matrix). Except for gender *(p = .896)*, mother's education *(p = .878)* and the ELO-sentences score *(p = .075)* at age 29 months, all the other 22 and 29 months measures were significantly positively correlated with the Productive Language Factor (see Table 6). Neither mother's education nor child's gender nor ELO-sentences score were included in subsequent analyses.

2.2.2.2.1. What is the predictive power of vocabulary measures at 22 months of age relative to lexical and grammatical skills and

Table 5
Component Matrix of the Language Productive factor extracted after the factorial analysis with language measures at 36 months for Study 1b.

Measure	Component 1
NEEL score 36 m.o.	.820
SRT score 36 m.o.	.696
LS-words score 36 m.o.	.855
LS-MLU score 36 m.o.	.840
Extraction Method: Principal Component Analysis	

Note: 1 components extracted ; NEEL, Nouvelles Epreuves d'Evaluation du Langage, subtest of productive vocabulary ; SRT, Sentence Repetition Task ; LS, Language Sample.

Table 6
Correlations between all predictor variables and Productive Language Factor at 36 months in Study 1b.

Measure	Productive Langage Factor W4
Maternal Education	.022
Sex	.019
IFDC-production score 22 m.o.	.386**
CCT score 22 m.o.	.426**
IFDC-production score 29 m.o.	.688**
ELO-words score 29 m.o.	.422**
ELO-sentences score 29 m.o.	.256
NRT score 29 m.o.	.626**
LS-words score 29 m.o.	.614**
LS-MLU score 29 m.o.	.570**

Note: IFDC, Inventaire Français du Développement Communicatif; CCT, Computerised Comprehension Task; ELO, Evaluation du Langage Oral, subtests of receptive skills; NRT, Nonword Repetion Task; LS, Language Sample.
*p < .05.
** p < .01.

phonological short-term memory at 29 months of age to predict productive language abilities at 36 months?. A hierarchical multiple linear regression was performed with the Productive Language Factor as a dependent variable. We entered both 22 months variables (IFDC-production score and CCT score) in Step 1, and all 29 months variables (IFDC-production score, ELO-words, NRT score, LS-words score, and LS-MLU score) in Step 2 using the stepwise method. Tolerance values were above .412 at every step indicating that independent variables were sufficiently different to fit into the models.

Results indicate that a total of 65.0% of the variance of the Productive Language Factor at 36 months, $F(648) = 15.888$, $p < .001$, could be accounted for by two variables at 22 months (the number of words understood on the CCT, and the number words produced on the IFDC), and by four additional variables at 29 months (the number of words produced on the IFDC, the number of different words produced during the Free Play, the number of phonemes repeated in the NWT and the number of words comprehended in the ELO task). See Table 7 for all regression parameters.

As in Study 1a, we contrasted the six models using the AIC approach (Posada & Buckley, 2004). The approach yielded AIC of 134.77, 136.68, 113.14, 121.43, 120.25 and 134.96 from Models 1 to 6, respectively. The observed difference in AIC scores show us that Model 3 (parent-reported expressive vocabulary on the IFDC) provides overall the best fit to predict the Productive Language Factor at 36 months of age.

2.2.2.2.2. What is the predictive power of vocabulary measures at 22 months of age relative to lexical and grammatical skills and phonological short-term memory at 29 months of age to predict lexical comprehension abilities at 36 months?. To assess predictors of lexical comprehension, we also performed a hierarchical multiple regression with the EVIP score, the only dependent variable that measures comprehension at W4. We entered all 22 months variables (IFDC-production and CCT scores) in Step 1 and all 29 months variables (IFDC-production, ELO-words, NRT, LS-words and LS-MLU scores) in Step 2 using the stepwise method. The results show that the CCT score at 22 months, $F\Delta(1,47) = 17.878$, $p < .001$, $R^{2\Delta} = .260$, and the ELO-words score at 29 months, $F\Delta(1,46) = 6.924$, $p = .012$, $R^{2\Delta} = .095$, explained a total of 37.0% of the receptive vocabulary score at 36 months, $F(2,48) = 13.528$, $p < .001$. See Table 8 for all regression parameters. Tolerance values were above .760 at every step indicating that independent variables were sufficiently different to fit into the models.

The AIC approach yielded values of 123.2 and 123.3 for the CCT score (W2) and ELO-words score (W3) models, respectively. The lack of difference observed in AIC scores suggests that both models provide an equally good fit to the data.

2.2.3. Discussion

The main aim of the second part of the study was to assess the predictive power of various tests of language acquisition measured at two time points (22 and 29 months) in predicting language production and lexical comprehension abilities at 36 months of age.

2.2.3.1. Predictors of lexical comprehension evaluated at 36 months. Our results show that 37.0% of variance in lexical comprehension evaluated at 36 months can be explained by two direct measures assessing receptive vocabulary at younger ages. At age 22 months, the direct and decontextualized comprehension task (CCT) predicts 26% of the total variance of receptive vocabulary at age 36 months. When the standardized comprehension task (ELO-words) evaluated at age 29 months is added into the model, it accounts for an additional 9.5%. Both models have very similar AIC scores, which means that both are equally powerful to predict lexical comprehension at 36 months. However, it should be noted that the CCT is a more distant predictor (22 months) than the ELO-words (29 months), and that despite this temporal gap, it is more effective than the ELO score in predicting receptive language outcome at 36 months, which clearly enhances the interest of the CCT as an early predictor of later comprehension. It thus appears that lexical comprehension at 3 years is predicted only by earlier measures of lexical comprehension, which suggests that the prediction of this dimension is both modality specific and language level specific. Although the material is different (pictures presented on a computer screen versus on paper), it should be noted that the task requirements are similar between the predictors (CCT and ELO) and the

Table 7
Hierarchical regression parameters for models predicting Language Productive Factor at 36 months from predictors at 22 and 29 months (N = 49) in Study 1b.

	Model 1						Model 2						Model 3		
Measure	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	F^{Δ}	*SE*
	.181					**.002**	**.231**					**.001**	**.467**		
CCT-word's comprehension 22 m.o.		**.164**	10.414	.131	.426	**.002**				.130	.349	**.011**			.115
IFDC-word's production 22 m.o.								**.082**	5.094	.127	.296	**.029**			.133
IFDC-word's production 29 m.o.														21.310	.156
LS-NDword's production 29 m.o.															
NRT-phonemes 29 m.o.															
ELO-word's comprehension 29 m.o.															

	Model 3		Model 4						Model 5						Model 6					
Measure	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	p
		.000	**.560**					**.000**	**.606**					**.000**	**.650**					**.000**
CCT-word's comprehension 22 m.o.	.162	.171				.105	.154	.153				.104	.078	.457				.112	−.062	.587
IFDC-word's production 22 m.o.	−.086	.534				.121	−.112	.374				.119	−.198	.115				.115	−.269	**.029**
IFDC-word's production 29 m.o.	.674	**.000**				.151	.513	**.001**				.148	.424	**.004**				.142	.362	**.009**
LS-NDword's production 29 m.o.				**.097**	10.579	.114	.361	**.002**				.109	.333	**.003**				.103	.308	**.004**
NRT-phonemes 29 m.o.										.051	6.174	.133	.311	**.017**				.135	.431	**.002**
ELO-word's comprehension 29 m.o.																**.047**	6.420	.105	.277	**.015**

Note 1 : CCT, Computerised Comprehension Task; IFDC, Inventaire Français du Développement Communicatif; LS, Language Sample ; NRT, Nonword Repetition Task ; ELO, Evaluation du Langage Oral, subtests of receptive skills.
Note 2: df 1,2 from (1;47) for Model 1 through (1;42) for Model 6.

Table 8
Hierarchical regression parameters for models predicting the EVIP score at 36 months (N = 49) in Study 1b.

	Model 1						Model 2					
Measure	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*	R^2	$R^{2\Delta}$	F^{Δ}	*SE*	β	*p*
	.276					**.000**	**.370**					**.000**
CCT score 22 m.o.		**.260**	11.827	0.116	0.525	**.000**				.126	.352	**.012**
ELO-words score 29 m.o.								**.095**	13.528	.122	.353	**.012**

Note: CCT, Computerised Comprehension Task; ELO, Evaluation du Langage Oral, subtests of receptive skills.

outcome measure (EVIP), which are all pointing to pictures tasks.

2.2.3.2. Predictors of production skills evaluated at 36 months. Turning now to the predictors of language production, our results indicate that 65.0% of the variance of the Productive Language Factor, which encompasses lexical and grammatical aspects of both spontaneous language and standardised assessments at 3 years of age, is explained by 6 earlier language predictors. As in Study 1a, at the age of 22 months, expressive vocabulary reported by the parents and the decontextualized comprehension task explain respectively 8.2% and 16.4% of the variance of the Productive Language Factor. Consistent with Study 1a, the amount of variance was superior for the CCT at this time-point relative to parent report. These results confirm that before 2 years of age, the strongest predictor of later language production is the direct measure of lexical comprehension (Friend et al., 2012; Friend, Smolak, Patrucco-Nanchen et al., 2018).

However, by 29 months, parent reported vocabulary production (IFDC) accounts by itself for an additional 23.7% of the variance of Productive Language Factor in fact, given the common variance that the IFDC shares with the variables measured at 22 months, it suppresses the effect of these earlier variables, which do not appear significant any longer.

The major contribution made by the IFDC-word production index at age 29 months was rather unexpected given that previous studies (Duff, Reen et al., 2015; Henrichs et al., 2011; Lee, 2010) reported that it accounted only for a small proportion of outcome variance. It is the only measure that is significant in Model 3, which has the best fit to the data according to the AIC analyses. This result thus confirms that at the age of two and one half years, parent reports of vocabulary production, constitute a very reliable predictor of later language skills when these are broadly assessed (Bates, Dale, & Thal, 1995; Bauer, Goldfield, & Reznick, 2002; Duff, Reen et al., 2015, Duff, Nation et al., 2015; Kemp et al., 2017; Marchman & Fernald, 2008a, 2008b; Morgan et al., 2015). Our results are in line with those of others in indicating that up to the age of 2 years, parent reports of vocabulary production account for a limited portion of variance in language outcome at 3 years of age (Duff, Reen et al., 2015, Duff, Nation et al., 2015; Ghassabian et al., 2013; Henrichs et al., 2011; Marchman & Fernald, 2008a, 2008b; Peyre et al., 2014), but that after the age of two and one half years, parent-reported vocabulary production is a strong predictor of language production at 36 months (Kemp et al., 2017; Marchman & Fernald, 2008a, 2008b; Morgan et al., 2015). The increased predictive power of the parental report measure according to the age of children are in line with the notion of stability addressed by Bornstein et al. (2016), in the sense that it may be difficult for more distal predictors to account for more or similar variance than more proximal ones (Bornstein, Putnick, & Esposito, 2017).

The other measures tested at 29 months, lexical diversity, non-word repetition accuracy and direct vocabulary comprehension (ELO-words), each account for 4–10% of additional variance of the Productive Language Factor, but adding these variables resulted in a decline of the fitness of the models. This suggests that their contribution is too limited to add useful information to the prediction. As for the direct measure of lexical production at 29 months, lexical diversity correlates quite strongly with the IFDC production score, suggesting a sizeable overlap between the two measures. Parental reports most likely capture the lexical production skills of their child in various settings and can therefore be considered more representative of the child's actual vocabulary size than a sample collected in a 15 min free play situation involving a limited set of toys and objects, eliciting a restricted number of words.

On the basis of previous results showing the link between proficiency in non-word repetition and lexical and grammatical production skills (Adams & Gathercole, 1995; Botting & Conti-Ramsden, 2001; Chiat & Roy, 2007; Dispaldro, Deevy, Altoé, Benelli, & Leonard, 2011; Gathercole, Willis, Emslie, & Baddeley, 1992; Gathercole & Adams, 1993, 2000; Stokes & Klee, 2009), we hypothesized that non-word repetition would have a large unique contribution to the children's language production outcome. In a way, our results extend these findings to a younger age group, but the fact that this contribution is limited (5.1%) can also be taken to indicate that this measure may not be sufficiently reliable at two and half years.

2.2.3.3. General discussion of study 1b. Overall, our results extend previous findings in two ways. On the one hand, our study shows that before 2 years of age, the most reliable predictor of lexical comprehension and language production at 3 years is the direct measure of lexical comprehension (Friend, Smolak, Patrucco-Nanchen et al., 2018, 2012). On the other hand, our models with 2 language predictors assessed at 22 months, and 4 language predictors assessed at 29 months, explained respectively 37.0% of the variance in receptive language, and 65% of the variance in expressive language at 3 years. These values are larger than those reported by Reilly et al. (2010) on the language ability of over 1500 Australian children at age 4 years, whose vocabulary skills had been measured at 2 years. Their multifactorial model with 13 predictors, explained 23.6% of the variance in receptive language, and 30.1% of the variance in expressive language. Furthermore, the CCT assessed at 22-months accounts respectively for 27.3% and 18.1% of the variance of receptive and expressive language outcome at 3 years of age, which again is much higher than the values typically reported using lexical comprehension and production questionnaires (Duff, Reen et al., 2015, Duff, Nation et al., 2015; Henrichs et al., 2011; Peyre et al., 2014; Reilly et al., 2010; Zambrana et al., 2014). Our results thus confirm that parent report of lexical knowledge collected before the age of 2 years is an unstable predictor of later language (Duff, Reen et al., 2015, Duff, Nation et al., 2015; Fernald & Marchman, 2012; Henrichs et al., 2011; Morgan et al., 2015; Peyre et al., 2014). However, contrary to Duff, Nation et al. (2015) claim, our results suggest that language skills can be reliably measured before 3 years of age and that the strongest predictor of language at age 3 varies as a function of the age at which prediction is undertaken. Before 24 months of age, a direct assessment of early comprehension provides the best fit to the data whereas, by 29 months, the parent report of vocabulary production is the most reliable measure, even though a set of language measures evaluating phonological, lexical and grammatical skills allows to increase the part of explained variance of the language skills of pre-schoolers –but at the cost of model fitness and economy.

3. General discussion

Early language development is considered critical for children's adjustment in school, for social adaptation and for later educational achievement. Despite the role of children's receptive skills as a foundation for later productive word use, receptive language skills have received surprisingly little attention. Instead, the study of language development has been centred on language production, which is easier to assess (for a review, Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008). Several recent studies show that poor language comprehension at 18 months, assessed by parent report, is the most reliable early predictor of later expressive vocabulary delay at the group level (Ellis Weismer, 2007; Henrichs et al., 2011; Zambrana et al., 2014) and that children with early delays in language comprehension in addition to language production are more likely to display persistent language delay (Bishop et al., 2012; Ellis Weismer, 2007; Henrichs et al., 2011).

The results of the current study, together with those of other studies, suggest that up to the age of 2, although parent reports of lexical comprehension and/or production account for a portion of variance in later receptive, productive or general language outcome (Henrichs et al., 2011; Reilly et al., 2010; Zambrana et al., 2014), they have less predictive validity (Duff, Reen et al., 2015, Duff, Nation et al., 2015; Feldman et al., 2000; Fernald & Marchman, 2012; Ghassabian et al., 2014; Henrichs et al., 2011) than a direct measure of early comprehension. By contrast, after age 2, parent report of vocabulary production is the strongest predictor of later language production skills, broadly assessed (Kemp et al., 2017; Marchman & Fernald, 2008a; Morgan et al., 2015) six months later. Our results extend previous findings in four ways. First, our study confirms the strong relation between early lexical comprehension and later language production abilities by showing that, before two years of age, the CCT administered at 22 months is the most reliable predictor of general language outcome at 29 months and language production at 36 months (Ellis Weismer, 2007; Friend et al. [ab]'s, 2012; 2018b; Henrichs et al., 2011; Zambrana et al., 2014). Secondly, we find that the CCT administered at 22 months accounts for significant unique variance in language comprehension at 36 months and is comparable in strength than the ELO direct assessment of comprehension administered at 29 months of age. Thirdly, the unique variance in receptive, expressive, and broad language outcomes accounted for by the CCT amounts to values ranging from 16.4 to 43.2%, which is three to eight times larger than has been reported for parent report measures (Duff, Reen et al., 2015, Duff, Nation et al., 2015; Henrichs et al., 2011; Peyre et al., 2014; Reilly et al., 2010; Zambrana et al., 2014). Finally, contrary to Duff, Reen et al.'s (2015), Duff, Nation et al.'s (2015) proposition, our results suggest that language skills can be reliably measured before 3 years of age. Our findings are in line with those of Conti-Ramsden and Durkin (2012), according to whom "the assessment of language skills in preschool children should involve an evaluation of both expressive and receptive skills, should include an evaluation of more than one dimension of language and if possible also include a measure of phonological short-term memory abilities" (p.396).

4. Conclusion and future directions

In sum, we find that children's vocabulary can be measured with predictive validity to developmental outcomes as early as the second year of life. Further, we find that the age at which an assessment is performed affects its ability to predict downstream outcomes. For example, early comprehension at 22 months of age is a strong predictor of general language abilities 6 months later whereas parent reported production at 29 months of age is a strong predictor of general language *production* skills at 36 months. Notably, parent-reported production at 29 months does not predict *comprehension* at 36 months, the latter being better predicted by direct measures of comprehension. Thus, there appear to be issues of both timing and specificity (comprehension versus production versus general language ability) at play in the early prediction of the course of acquisition. This suggests that it is essential to employ complementary measures in assessing early language and further, that this can be accomplished with a rather small set of measures as long as some of these measures are direct measures of early vocabulary comprehension and production.

There are several limitations to the present study. Although we were able to show the power of an early direct lexical comprehension task administered at 22 months to predict later language comprehension and production abilities in French-speaking children, it remains unknown whether such a task could be employed cross-linguistically, and in multilingual samples. The small sample size is another limitation of this study and did not allow us to do more powerful statistical analyses and to conduct relative importance analyses to evaluate the predictors independently. As a result, this experiment should be replicated with larger samples of children. Furthermore, the limited range of SES exhibited by our family sample also constitutes a shortcoming. Finally, given that several studies showed that most children with both a receptive and an expressive lexical delay at 2 years of age display oral language and academic difficulties, and are at risk of subsequently developing learning disorders as well as cognitive, behavioural and psychiatric difficulties, it would be useful to scale up assessment with the CCT to larger groups of toddlers between 18 and 24 months to allow a normalization of the task. It would also be interesting to examine whether the speed of processing measured with the CCT could be considered as an additional predictor of subsequent language outcome. The use of the CCT and IFDC as complementary measures would allow clinicians to direct children with receptive and expressive delay more quickly to speech and language therapists and enable them to monitor the progress of these children, support parents and implement early intervention aiming to limit the risks mentioned above and reduce inequities at school entry.

Funding

This research was supported by the National Institutes of Health [NICHD under award #R01HD468058] and does not necessarily represent the views of the National Institutes of Health.

Declaration of Competing Interest

Tamara Patrucco-Nanchen, Margaret Friend, Diane Poulin-Dubois and Pascal Zesiger declare that they have no conflict of interest.

Acknowledgments

We wish to thank Laura Alaria, Julie Berclaz, Tess Cornélius, Anne-Sophie Daviet, Pauline Gavillet, Mélissa Icart, Elise Okan, Valentine Pfister, Liliana Silva and Matthieu Vétois for assistance in data collection and coding. We also gratefully acknowledge all of our participants and families, Audrey Buerki for his statistical advice and Margaret Kehoe for the careful reading of this manuscript.

Appendix A

Table A1
French Non-Word Repetition Task based on Hoff et al. (2008).

		Non-words	Phonological production	Construction of the non-words
training	**1**	gotte	/gɔt/	botte
training	**2**	dambe	/dãb/	jambe
	1	dulle	/dyl/	bulle
	2	bame	/bam/	dame
	3	soupe	/ʀup/	soupe
	4	sobe	/sɔb/	robe
	5	mallon	/malõ/	matin/maman - ballon
	6	gabou	/gabu/	garage - hibou
	7	souteur	/suτœʀ/	souris - docteur/facteur/tracteur
	8	fonette	/fɔnετ/	forêt - lunettes
	9	noupireuil	/nupiʀœj/	nourriture - papillon - écureuil
	10	pylécot	/pileko/	pyjama - éléphant - haricot
	11	téfilon	/τefilõ/	téléphone - confiture - pantalon
	12	cicomal	/sikɔmal/	cinéma - crocodile - animal

Appendix B

Table B1
French Sentence Repetition Task based on Devescovi & Caselli, 2007.

	Simple sentences with copula	#syll.	#words	#morph.
1	Le chat est joli. *The cat is cute.*	5	4	5
2	La voiture est rouge. *The car is red.*	5	4	5
3	La souris est petite. *The mouse is small.*	6	4	6
	Simple sentences with one argument			
	Singular			
4	Le chien mange. *The dog is eating.*	3	3	4
5	La fille danse. *The girl is dancing.*	3	3	4
6	Le lion court. *The lion is running.*	3	3	4
	Plural			
7	Les bébés dorment. *The babies are sleeping.*	4	3	5
8	Les enfants lisent. *The children are reading.*	4	3	5
9	Les chevaux boivent. *The horses are drinking.*	4	3	6
	Sentences with one argument and one modifier			
10	Le lapin saute loin. *The rabbit is jumping far.*	5	4	5
11	La fleur pousse vite. *The flower is growing fast.*	4	4	5
12	Les garçons chantent encore. *The boys are still singing.*	6	4	6
	Simple sentences with two arguments			
13	Le singe prend la banane. *The monkey is taking the banana.*	6	5	6
14	La vache regarde le train. *The cow is looking at the train.*	6	5	6
15	La dame ouvre la fenêtre. *The woman is opening the window.*	6	5	6
	Simple sentences with two arguments and a simple preposition			
16	Le cochon va à la plage. *The pig is going to the beach.*	7	6	7
17	La poupée tombe sous la table. *The doll is falling under the table.*	7	6	7
18	Le poisson nage dans l'eau. *The fish is swimming in the water.*	6	6	7
	Sentences with two arguments and one modifier			
19	L'oiseau porte le grand papillon. *The bird is carrying the big butterfly.*	8	6	8
20	L'ours conduit la moto bleue. *The bear is driving the blue motorbike.*	7	6	8
21	La chèvre prend la belle carotte. *The goat is taking the beautiful carrot.*	7	6	8
	Simple sentences with three arguments and a simple preposition			
22	Lola donne la main à Hugo. *Lola is giving the hand to Hugo.*	8	6	7
23	Manon met le fromage sur la table. *Manon is putting the cheese on the table.*	9	7	8
24	Noah lave le camion dans le garage. *Noah is washing the truck in the garage.*	10	7	8
	Simple sentences with three arguments and a compound preposition			
25	Chloé apporte le pain au canard. *Chloé is bringing (the) bread to the duck.*	9	6	8
26	Louis ferme la porte du salon. *Louis is closing the door of the living room.*	8	6	8
27	Théo lit le livre aux enfants. *Théo is reading the book to the children.*	8	6	8

References

Adams, A.-M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech Language and Hearing Research, 38*(2), 403–414. https://doi.org/10.1044/jshr.3802.403.

Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders, 35*, 95–116.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher, & B. Mac Whinney (Eds.). *The handbook of child language* (pp. 96–151). Oxford: Blackwell.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., et al. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*(01), 85–123. https://doi.org/10.1017/S0305000900008680.

Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analysing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics, 23*(03), 313–335. https://doi.org/10.1017/S0142716402003016.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Non-word repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37*, 391–403.

Bishop, D. V., Holt, G., Line, E., McDonald, D., McDonald, S., & Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: A longitudinal study of precursors of specific language impairment. *Journal of Neurodevelopmental Disorders, 4*(1), 3. https://doi.org/10.1186/1866-1955-4-3.

Bornstein, M. H., Hahn, C.-S., & Putnick, D. L. (2016). Long-term stability of core language skill in children with contrasting language skills. *Developmental Psychology, 52*(5), 704–716. https://doi.org/10.1037/dev0000111.

Bornstein, M. H., Putnick, D. L., & Esposito, G. (2017). Continuity and stability in development. *Child Development Perspectives, 11*(2), 113–119. https://doi.org/10.1111/cdep.12221.

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 36*(4), 421–432. https://doi.org/10.1080/13682820110074971.

Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *The British Journal of Developmental Psychology, 26*(2), 281–300. https://doi.org/10.1348/026151007X235891.

Chevrie-Muller, C., & Plaza, M. (2001). *Nouvelles épreuves pour l'examen du langage.* ECPA.

Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech Language and Hearing Research, 50*(2), 429–443. https://doi.org/10.1044/1092-4388(2007/030).

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 43*(1), 1–40. https://doi.org/10.1080/13682820601116485.

Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology Review, 22*(4), 384–401. https://doi.org/10.1007/s11065-012-9208-z.

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The language exposure assessment tool: Quantifying language exposure in infants and children. *Journal of Speech Language and Hearing Research, 59*(6), 1346–1356. https://doi.org/10.1044/2016_JSLHR-L-15-0234.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*(4), 361–389. https://doi.org/10.1080/13682820701546854.

Devescovi, A., & Caselli, C. M. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language & Communication Disorders, 42*(2), 187–208. https://doi.org/10.1080/13682820601030686.

Dispaldro, M., Deevy, P., Altoé, G., Benelli, B., & Leonard, L. B. (2011). A crosslinguistic study of real-word and nonword repetition as predictors of grammatical competence in children with typical language development. *International Journal of Language & Communication Disorders, 46*(5), 564–578.

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.

Duff, F. J., Nation, K., Plunkett, K., & Bishop, D. (2015). Early prediction of language and literacy problems: is 18 months too early? *PeerJ, 3*, e1098. https://doi.org/10.7717/peerj.1098.

Dunn, L. M., Dunn, L. M., & Theriault-Whalen, C. (1993). *EVIP: échelle de vocabulaire en images Peabody.* Pearson Canada Assessment.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test.* American Guidance Service.

Ellis Weismer, S. (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.). *The influence of developmental perspectives on research and practice in communication disorders: A festschrift for Robin S. Chapman* (pp. 83–101). Mahwah, NJ: Lawrence Erlbaum Associates.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 856–868.

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.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur communicative development inventories: User's guide and technical manual.* San Diego: Singular Publishing.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers: Lexical processing and vocabulary growth. *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), 98–116. https://doi.org/10.1037/0012-1649.42.1.98.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228–231.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing, 44*, 97–135.

Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods Instruments & Computers, 35*(2), 302–309.

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*, 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., & Zesiger, P. (2011). Une réplication systématique des propriétés psychométriques du Computerized Comprehension Task dans trois langues. *Enfance, 3*, 329–344. https://doi.org/10.4074/S0013754511003053.

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), 1317–1333. https://doi.org/10.1037/dev0000514.

Friend, M., Smolak, E., Patrucco-Nanchen, T., Poulin-Dubois, D., & Zesiger, P. (2018). Language status at age 3: Group and individual prediction from vocabulary comprehension in the second year. *Developmental Psychology*. https://doi.org/10.1037/dev0000617 October 25 Advance online publication.

Fujiki, M., Brinton, B., Isaacson, T., & Summers, C. (2001). Social behaviors of children with language impairment on the playground: A pilot study. *Language, Speech, and Hearing Services in Schools, 32*(2), 101–113.

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*(4), 770–778.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school \ears: A longitudinal study. *Developmental Psychology, 28*(5), 887–898.

Ghassabian, A., Rescorla, L., Henrichs, J., Jaddoe, V. W., Verhulst, F. C., & Tiemeier, H. (2014). Early lexical development and risk of verbal and nonverbal cognitive delay at school age. *Acta Paediatrica, 103*(1), 70–80. https://doi.org/10.1111/apa.12449.

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*(01), 23–45.

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.

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). *Using language sample databases. language, speech, and hearing services in schools*.

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.

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., et al. (2011). Examining continuity of early expressive vocabulary development: The generation r study. *Journal of Speech Language and Hearing Research, 54*(3), 854–869. https://doi.org/10.1044/1092-4388(2010/09-0255).

Hilbe, J. M. (2011). *Negative binomial regression.* Cambridge University Presshttps://books.google.com/books?id=DDxEGQuqkJoC&pg=PA43&source=gbs_toc_r&cad=4#v=snippet&q=aic&f=false.

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). Cambridge, MA, US: The MIT Press.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*(1), 5588. https://doi.org/10.1016/j.dr.2005.11.002.

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language, 35*(04), 903–916. https://doi.org/10.1017/S0305000908008751.

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*(04), 701–724. https://doi.org/10.1017/S0305000907008124.

Howell, D. C. (1997). *Statistical methods for psychology.* Belmont, California: Wadsworth, Fourth.

Irwin, J. R., Carter, A. S., & Briggs-Gowan, M. J. (2002). The social-emotional development of "late-talking" toddlers. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(11), 1324–1332.

Kemp, N., Scott, J., Bernhardt, B. M., Johnson, C. E., Siegel, L. S., & Werker, J. F. (2017). Minimal pair word learning and vocabulary size: Links with later language skills. *Applied Psycholinguistics, 38*(2), 289–314.

Kern, S. (1999). *Inventaire Français du Développement Communicatif chez le nourrisson: Mots et gestes.* Lyon, France: Laboratoire Dynamique du Langage. Institut des Sciences de l'Homme.

Kern, S., & Gayraud, F. (2010). *Inventaire Français du Développement Communicatif: 8/30 mois.* Les Editions de la Cigale.

Khomsi, A. (2001). *ELO: évaluation du langage oral.* ECPA, Les Editions du Centre de psychologie appliquée.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine and Child Neurology, 42*(03), 190–200.

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health, 13*(4), 198–206. https://doi.org/10.1111/j.1475-3588.2008.00503.x.

Law, J., Rush, R., Parsons, S., & Schoon, I. (2013). The relationship between gender, receptive vocabulary, and literacy from school entry through to adulthood. *International Journal of Speech-language Pathology, 15*(4), 407–415. https://doi.org/10.3109/17549507.2012.721897.

Lee, J. (2010). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(01), 69–92. https://doi.org/10.1017/S0142716410000299.

Marchman, V. A., & Fernald, A. (2008a). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x.

Marchman, V. A., & Fernald, A. (2008b). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x.

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.). *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–124). Multilingual matters.

Miller, J. F., Iglesias, A., & Nockerts, A. (2012). *SALT: Systematic analysis of language transcripts (research version) [computer software].* Middleton, WI: SALT Software.

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.

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish–English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology, 108*(1), 113–125. https://doi.org/10.1016/j.jecp.2010.07.011.

Peyre, H., Bernard, J. Y., Forhan, A., Charles, M.-A., De Agostini, M., Heude, B., et al. (2014). Predicting changes in language skills between 2 and 3 years in the EDEN mother–child cohort. *PeerJ, 2*, e335. https://doi.org/10.7717/peerj.335.

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: what does the task measure? *International Journal of Language & Communication Disorders, 50*(1), 106–118.

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in Phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology, 53*(5), 793–808. https://doi.org/10.1080/10635150490522304.

Redmond, S. M., Thompson, H. L., & Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research, 54*(1), 99–117. https://doi.org/10.1044/1092-4388(2010/10-0010).

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., et al. (2010). Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics, 126*(6), e1530–e1537. https://doi.org/10.1542/peds.2010-0254.

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech Language and Hearing Research, 52*(1), 16–30.

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141–150. https://doi.org/10.1002/ddrr.1108.

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech Language and Hearing Research, 51*(2), 394–407.

Schoon, I., Parsons, S., Rush, R., & Law, J. (2010). Childhood language skills and adult literacy: A 29-Year follow-up study. *Pediatrics, 125*(3), e459–e466. https://doi.org/10.1542/peds.2008-2111.

Snowling, J., Adams, J. W., Bishop, D. V. M., & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International Journal of Language & Communication Disorders, 36*(2), 173–183.

Snowling, M. J., Bishop, D. V. M., Stothard, S. E., Chipchase, B., & Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment: Psychosocial outcomes of LI children. *Journal of Child Psychology and Psychiatry, 47*(8), 759–765. https://doi.org/10.1111/j.1469-7610.2006.01631.x.

Stokes, S. F., & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry, 50*(4), 498–505. https://doi.org/10.1111/j.1469-7610.2008.01991.x.

Styles, S., & Plunkett, K. (2008). 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*(04), 895–908. https://doi.org/10.1017/S0305000908009264.

Thordardottir, E. T. (2005). Early lexical and syntactic development in Quebec French and English: Implications for cross-linguistic and bilingual assessment. *International Journal of Language & Communication Disorders, 40*(3), 243–278. https://doi.org/10.1080/13682820410001729655.

Trudeau, N. (2007). *Evaluation du développement linguistique français. Encyclopédie sur le développement des jeunes*. London, ON: Réseau canadien de recherche sur le langage et l'alphabétisation. Extrait de1–8. [30.04.2019] http://www.literacyencyclopedia.ca/pdfs/topic.php?topId=231&fr=true.

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.

Zambrana, I. M., Pons, F., Eadie, P., & Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset: Trajectories of language delay from age 3 to 5. *International Journal of Language & Communication Disorders, 49*(3), 304–316. https://doi.org/10.1111/1460-6984.12073.

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development, 6*(2), 179–208. https://doi.org/10.1207/s15327647jcd0602_2.