Social Education Research
http://ojs.wiserpub.com/index.php/SER/

Review

Tools and Teaching Strategies for Vocabulary Assessment and Instruction: A Review

E. Dujardin[1,2*], P. Auphan[1,2], N. Bailloud[1,3], J. Ecalle[1], A. Magnan[1]

[1]Laboratory EMC, EMC (EA 3085)-MSH LSE (USR CNRS 2005)-University Lyon 2 and LabEx CORTEX ANR-11-LABX-0042, University Lyon, France
[2]Association Agir Pour l'Ecole-Paris, France
[3]Canopé-Lyon, France
E-mail: emilie.dujardin@univ-lyon2.fr

Received: 16 July 2021; **Revised:** 17 November 2021; **Accepted:** 20 November 2021

Abstract: Vocabulary is a complex intermediate component between oral and written language, which the influence on associated skills and general language abilities (e.g., decoding processing, comprehension) has been largely studied in children, suggesting its important role in literacy. The main aim of this article is to review some questions on vocabulary assessment and stimulation in children and identify the advantage of new technologies for evaluating and training vocabulary. It seems necessary to give the importance of vocabulary in language and literacy development, and the heterogeneity of vocabulary acquisition, depending on preschool word exposure (e.g., familial environment). After a state-of-the-art of conceptualization of vocabulary's notion, we revisit assumptions on vocabulary assessment and instruction indicating the main existing tools. This review lies in the attempt to enhance perspectives for new valid and effective tools using digital technologies.

***Keywords*:** vocabulary, children, assessment, digital tool, teaching

1. Introduction

Vocabulary research has greatly expanded over the last three decades. Vocabulary is a multidimensional intermediate component between oral and written language (i.e. from decoding to comprehension). Numerous studies have demonstrated the impact of vocabulary on associated skills, such as general language abilities (Brinchmann et al., 2015), decoding processing (Tunmer & Chapman, 2012), comprehension (Cain & Oakhill, 2014; Ouellette & Beers, 2010; Quinn et al., 2015). Recent research cites developmental influences to explain the weight of vocabulary variation on these associated skills. When one starts to learn to read, vocabulary seems to contribute to decoding to consolidate the links among the three levels of word representation (i.e., orthographic, phonological, and semantic), but once these links are sufficiently consolidated with the acquisition of a high-quality lexicon, the link between vocabulary and decoding apparently disappears (Chiu, 2018; Massonnié et al., 2019) and vocabulary would then predict comprehension performance (Ouellette & Beers, 2010; Tilstra et al., 2009). Moreover, the link between vocabulary and associated skills (i.e., language and literacy skills) would be bi-directional. For example, vocabulary in young French students from grade 2 seemingly predicts decoding and reading performance and conversely, performance in decoding would predict

DOI: https://doi.org/10.37256/ser.3120221044

vocabulary performance (Potocki et al., 2016; Verhoeven et al., 2011). Further research is required to understand the role (causal or not (Quinn et al., 2015)), direct influence (Tunmer & Chapman, 2012) or indirect influence (Elwér et al., 2013; Ouellette & Beers, 2010), and the weight of vocabulary on associated skills. Despite the ongoing questions on the link between vocabulary and the other associated skills, there is currently a consensus on the importance of vocabulary in reading (Braze et al., 2016; Colenbrander et al., 2016; Perfetti et al., 2010; Oakhill et al., 2019) and it is essential to define what vocabulary corresponds to.

This article aims to provide a review of questions on the assessment and instruction of vocabulary in children, by revisiting assumptions on vocabulary assessment and instruction. First, we provide an overview of assumptions on vocabulary assessment and instruction. Second, we present the main tools that exist in different languages. Our review has been led to answer the following main research question: What is the advantage of new technologies for evaluating and training vocabulary. The added value of this review lies in the attempt to enhance perspectives for new valid and effective tools using digital technologies. Indeed, standardized presentation of items or precise recording of different kinds of measures are two characteristics of digital technologies that improve the validity and reliability of assessment (Terzis & Economides, 2011). Moreover, computer-based tools offer the opportunity to multiply and maximize learning sessions through autonomy in learning and appropriate and systematic feedback provided individually (Van der Kleij et al., 2015).

Without claiming to be a systematic review, this article reviews some questions on research on vocabulary assessment and instruction. It seems necessary to give the importance of vocabulary in language and literacy development and the heterogeneity of vocabulary acquisition that depends on preschool word exposure mainly in a family environment (Biemiller, 2006). Children exposed to language with poor vocabulary will experience more difficulties than their peers exposed to rich vocabulary in developing general language skills, learning to read and even learning the meanings of new words (Webb & Chang, 2015). Since the link goes both ways, children with language and literacy difficulties will then have greater difficulties in vocabulary learning than their peers who do not have language and literacy difficulties, thus illustrating the Matthew effect in vocabulary growth (Coyne et al., 2010).

Early vocabulary assessment and instruction, therefore, seem essential to reduce inequalities. To answer the main research question, after providing a brief definition of the complex concept of vocabulary, we present the main tools for assessing the various dimensions of vocabulary children across different languages. We then elaborate on the implications of vocabulary instruction and the associated recommendations.

In our current review, we have included research articles, book chapters, literature reviews, and meta-analyses, which cover a long period from 1982 to 2019 and provide insights into the evolution of vocabulary-related definitions, assessment and instruction. Thus, we identified 30 tests or subtests for vocabulary assessment or stimulation across different languages, suggesting that there is international interest in investigating the issue of vocabulary (see Appendix A for a summary of the referenced tests and subtests). Finally, to understand the evolution of vocabulary instruction, we identified 34 articles (corresponding to 43 studies; see Appendix B for more details).

2. What is vocabulary? An overview

2.1 *Which definition*?

The notion of "vocabulary" is expressed in different terms that must be clearly distinguished: the lexicon, word knowledge and vocabulary. The mental lexicon corresponds to the place (Ouellette, 2006) where word knowledge is stored. Word knowledge designates the information on the words (in particular orthographic, phonological, semantic, as well as general information associated with the target word (Perfetti & Hart, 2002). Finally, vocabulary refers to the grouping of knowledge presenting in the mental lexicon added to the processes that control the speed of access to that knowledge (Oakhill et al., 2012).

2.2 *Conceptualizing vocabulary: A multifaceted concept*

The concept of vocabulary is multifaceted. Broad and numerous conceptualizations have been described to characterize it to answer the following question: What does "knowing a word" mean? Four main conceptualizations

are used to specify vocabulary. The most common conceptualization is to make a distinction between the breadth (i.e., the number of words known by a subject) and the depth ((i.e., the quality of knowledge associated with the words) of vocabulary (Qian, 1999; Schmitt, 2014). Different types of associated knowledge can be considered on a continuum from superficial to deep. For example, knowledge can be syntagmatic (i.e., referring to properties of the object designated by the word) or paradigmatic (i.e., hierarchical vertical dimensions such as subordinate or superordinate levels) (Schwartz & Katzir, 2012). As described in Ordóñez et al. (2002), paradigmatic knowledge is for older subjects with higher school levels in relation with, for example, 'cognitive advances' or classroom instruction (Anglin et al., 1993; Snow, 1990).

Another distinction is made between declarative or procedural vocabulary knowledge (Nagy & Scott, 2000; Read, 2004). Declarative knowledge is currently described and is consciously and verbally accessible. Procedural knowledge rather refers to the implicit knowledge allowing children to appropriately and fluently pronounce and use words in context. Declarative knowledge could then imply "knowing a word" whereas procedural knowledge would imply "knowing how to use a word", representing a deeper level of vocabulary knowledge.

Three other main conceptualizations can be described. (a) The breadth and depth of vocabulary knowledge can be outranked from a network organization perspective (Meara & Wolter, 2004). Quality of word knowledge depends on the ability to link this word with a multitude of other words (breadth) in a coherent semantic organization (depth) such as collocations, synonyms or antonyms. A wider and better organized lexical network will facilitate rapid understanding and use of words in context. (b) Fluency is also a dimension that will represent vocabulary knowledge quality but there is no consensus on the way it should be conceptualized: As a part of depth (Segalowitz et al., 1998) or as an independent dimension in addition to size and depth (Daller et al., 2007). (c) Finally, Kieffer and Lesaux (2012) proposed another conceptualization sometimes seen as competing with the breadth versus depth distinction but that could also be thought of as being complementary. They distinguished between word-specific knowledge (i.e., linguistic knowledge of individual word meanings) and word-general knowledge (i.e., awareness of morphology). Word-general knowledge implies knowledge of "*the system by which complex words are formed from smaller meaningful units that contribute to their meanings and syntactic functions*" (Kieffer & Lesaux, 2012), for a description of morphological considerations and (Gardner, 2007). Word-general knowledge could eventually be related to strategies for learning vocabulary as presented in some studies such as Schmitt (2014).

Table 1. Conceptualizations on vocabulary depending on dimensions and task characteristics

Vocabulary dimensions	Task characteristics
Breadth (number of words known) *versus* Depth (quality of knowledge, from superficial-syntagmatic and/or declarative-to deep-paradigmatic and/or procedural-knowledge) (Schmitt, 2014)	Oral *versus* written (Nation, 2001)
Network organization (quantity and quality of word and concept links) (Meara & Wolter, 2004)	Receptive (passive) *versus* productive (active) (Nation, 2001)
Fluency (speed of access, between the flexibility of representation levels) (Daller et al., 2007)	Contextualized *versus* decontextualized (Read, 2000)
Word general (vocabulary breadth) *versus* word specific knowledge (words targeting a specific field) (Kieffer & Lesaux, 2012)	Discrete (vocabulary as an independent construct) *versus* embedded (vocabulary as part of a larger construct) (Coombe, 2011)
	Selective (vocabulary assessment only) *versus* comprehensive (vocabulary items in another cognitive assessment) (Pearson et al., 2007)

Note: See Appendix A for more details about vocabulary dimension assessment

But the vocabulary can also be characterized by the following distinctions that relate to task characteristics. Indeed, the task can imply different types of input and output such as oral as opposed to written, and receptive (passive) as opposed to productive (active) vocabulary (Nation, 2001). These two parameters enable a distinction to be made between four competencies described in the National Reading Panel (NICHD, 2000) (National Institute of Child Health and Human Development (NICHD)): Listening (oral and receptive), reading (written and receptive), speaking (oral and productive) and writing (written and productive). In general, children exhibit a larger receptive vocabulary than productive vocabulary (i.e., understand more words than they use) (Pearson et al., 2007). As described in Read (2000), a task can also include a word in context (contextualized) or presented in isolation (decontextualized). Moreover, a task can examine vocabulary as an independent construct (discrete vocabulary) or as a part of another cognitive construct such as reading comprehension for example (embedded vocabulary) (Coombe, 2011), or evaluate vocabulary for itself (selective) or in the case of more general tasks (comprehensive) (Pearson et al., 2007). Seeing Table 1 for a recap of vocabulary conceptualizations.

2.3 *Summary*

To go beyond the multiple and sometimes concurrent ways of conceptualizing vocabulary, the choice was made in this review to target the objective justifying the evaluation and training of vocabulary, for example, through the associated skills in which it participates. Indeed, knowing words is in itself a multiple concept involving the ability to: recognize and decode them rapidly (e.g., in breadth, decontextualized and discrete dimensions) that implies firstly linking the three levels of representation of the word described in the lexical quality hypothesis (Perfetti & Hart, 2002) and secondly the speed of access from one to the other (e.g., fluency and network organization dimensions). Knowing words requires understanding them in context (e.g., breadth and depth, contextualized, embedded, and receptive dimensions as well as word-general dimensions) in order to enhance the text comprehension processes and use them appropriately to express ideas (e.g., breadth and depth, contextualized, embedded, and expressive dimensions).

3. How to assess vocabulary? A literature review

3.1 *What research shows?*

The complexity of vocabulary implies complex assessment. One test is not enough to determine the vocabulary level because of the multiple facets of vocabulary. Different points need to be resolved when dealing with vocabulary assessment. First, which dimensions need to be assessed? Then, which test format should be used to examine vocabulary? Those questions are interrelated in the sense that the assessment test format will depend on the dimension to be assessed.

3.2 *Which dimensions need to be assessed?*

Beyond the distinction described, i.e. breadth as opposed to depth (Laufer & Goldstein, 2004; Qian & Schedl, 2004), selective as opposed to embedded (Coombe, 2011), or specific *versus* general word knowledge (Kieffer & Lesaux, 2012), vocabulary assessment is primarily divided into two parts: assessing vocabulary knowledge (i.e., type of knowledge about words) as opposed to assessing vocabulary learning (i.e., the ability to learn new words). The choice of dimension to be assessed will depend on the goal. For example, following an intervention, it could be interesting to assess specific word knowledge (adapted assessment depending on the material taught) associated with an assessment of general word knowledge (standardized test) in order to measure acquisition of targeted words taught and meta-competencies to learn new words (Tseng et al., 2006).

In recent reviews on vocabulary assessment, authors recommended assessing both breadth and depth of vocabulary as well as both word-specific learning and general vocabulary development (Hoffman et al. (2014) in early childhood) and, assessing target words in isolation, then in a context to measure whether they can be understood during reading and whether they possibly enable more general text comprehension (Kremmel & Schmitt, 2016).

3.3 *Which test format to examine vocabulary knowledge?*

A large variety of test formats are currently used depending on the dimension chosen to assess vocabulary (see Table 2 for a summary of the pros and cons of each test format). Moreover, the choice of test format will depend in part on the population to be assessed. For example, the production of definitions that are then scored by practitioners are better for younger children (<10 years old), since it appears as a more concrete task, and multiple-choice formats seem better for older (>10 years old) who prefer more abstract and analytic tasks (Read, 2004).

Table 2. Advantages and disadvantages of each test format

Test format	Description	Advantages	Disadvantages	References
Knowledge scale	Scale administered to test-takers in order to get their own thoughts about knowledge of targeted words	- Assesses different depths of word knowledge: from recognizing a word to being able to use it appropriately - Interesting for targeted words	- Does not take polysemy into account - Long to administer - Difficulties in scoring productive answers that measure more in-depth vocabulary knowledge	Dale (1965); Paribakht & Wesche (1993)
Multiple matching	Correctly paired items from two different list of items	- Rapid and easy to administer and score - Objective - Large number of items possibly administered in a short period of time - Widespread familiarity - Vocabulary size	- Discrete, context independent - Guessing behavior (can be diminished using "various number of correct answers") - Distractors choices	Gyllstad et al., (2015); Kremmel & Schmitt (2016)
Multiple choice	Selecting the one right answer to a question from distractors			
C-test	Completing a text with missing words	- Productive and in context	- Scoring: an answer can be correct but not match teachers expectation and targeted assessment	Harsch & Hartig (2016)
Yes/no test	Answering questions with yes or no	- Rapid and easy to administer - Interpreting hits and false alarm separately can increase reliability	- Low sensitivity, reliability and validity due to 50% chance of correct answers	Harsch & Hartig (2016); Hoffman et al. (2014)
Open-ended questions	Answering questions or generating definitions	- In-depth vocabulary assessment	- More subjective - Long to administer - Complex and long to score	Scalise & Gifford (2006)

Note: C-test = Cloze test

3.3.1 *Receptive size test formats*

Standardized vocabulary size tests (e.g. multiple choice or multiple matching) are the most common due to the objectivity, reliability and validity they offer. However, these kinds of tests have drawbacks. Firstly, standardized tests do not enable specific words taught during instruction to be assessed. Assessing word learning after instruction will require the test to be adapted to the material taught resulting in a decrease in validity and reliability (Hoffman et al., 2014; Watkins & DeThorne, 2000). Secondly, even if this enhanced objective scoring and reliability, these tests are based on a forced-choice format that can encourage guessing behavior leading to biased test results (Gyllstad et al., 2015). Nevertheless, guessing behavior can be limited using varying and unknown number of correct answers, and interpreted as indicators of learning or processing strategies by tracking test takers' behavior.

Another kind of forced-choice format is the Yes-No format. Despite the advantage of rapid and easy administration, this format gives a 50% chance of correct random responses, impacting the sensitivity of the task (Hoffman et al., 2014) and interpretation of results (Pellicer-Sánchez & Schmitt, 2012; Stubbe, 2012). However, interpreting hits and false alarm scores separately could increase reliability (Harsch & Hartig, 2016).

3.3.2 *Productive depth test formats*

Knowledge scales consist of self-assessment of vocabulary knowledge. They enable a distinction to be made between recognizing the word and being able to use it appropriately in context (Paribakht & Wesche, 1993). Therefore,

these tests are a good way to assess the depth of vocabulary and offer the opportunity to be easily adaptable to targeted taught words. However, they have two main drawbacks: Firstly, since different levels of vocabulary knowledge (from breadth to depth) are assessed item by item, the vocabulary scale is then long to administer. Secondly, the greater the vocabulary depth, the more difficult it is to score.

Cloze tests (or C-tests) consist of text with missing words to be found. They can be administered in multiple-choice or free answer formats. In the case of the free answer format, the advantage is that this is a productive task provided in context. However, this is also a disadvantage since the subject's answer may be correct, but may not be the one expected by the teacher (targeting words taught). Finally, open-ended questions favor an assessment of vocabulary depth, but are complex to score and can appear as being more subjective. Therefore, open-ended tasks cannot be standardized.

3.4 *Summary*

Assessing vocabulary requires using not only one test but a battery of tests to assess different dimensions (size, depth, fluency), using different cognitive parameters (oral or written, receptive or productive) and conditions (contextualized or decontextualized, discrete or embedded) as well as assessing specific words or general words and also vocabulary learning strategies. Another difficulty lies in producing "intermediate constraint" items (for a review of the different kinds of items depending on their constraint level, refer to Scalise & Gifford (2006)) that are not fully selective (like multiple-choice tasks) or constructed (like open-ended questions or essays). Indeed, items that only require selecting an answer offer the advantage of being easily scored and objective but can overestimate subjects' performances due to guessing behavior. Moreover, they usually give more information about vocabulary size than depth. Inversely, tasks that require building the answer from scratch are complex and more subjective to score but can provide more information on vocabulary depth and prevent guessing behavior.

4. Existing tools for assessing vocabulary

4.1 *General measures of the breadth of vocabulary*

The breadth of vocabulary is simple to conceptualize because it refers to the number of lexical items stored in the lexicon. Standardized measures of breadth usually test the connection between the form and meaning of words. The purpose of this section is not to provide an exhaustive list of all existing tests but to identify the most frequently used ones in different languages (see Appendix A).

Two measures of the breadth of an individual's vocabulary are identified in the literature Receptive vocabulary measures assess the connection between the form and the meaning of words, and expressive vocabulary measures assess the connection between the meaning and form of a word. The British Picture Vocabulary Scale (BPVS) (Dunn et al., 2009) and the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 2007) are the most common breadth measures. This assessment helps identify the receptive level of the child's language as well as learning difficulties. The PPVT has been largely translated in different languages, in French (Dunn et al., 1993); in Spanish (Dunn et al., 1986); in German (Rossbach et al., 2005). These are simple, easy-to-administer tests in which an examiner shows a set of four images and asks children to indicate which image best matches the word provided by the examiner. These tests can help identify a verbal delay even before the children start school.

The breadth of vocabulary is also related to expressive vocabulary with two common examples, the Expressive One-Word Picture Vocabulary Test (EOWPVT), in English (Martin & Brownell, 2010); in Spanish (Brownell, 2001); in French (Groupe coopératif en orthophonie-Région Laval, Laurentides Lanaudière, 1995) and The Expressive Vocabulary Test (EVT), in English (Williams & Pearson Education, 2018). In these tests, the examiner presents a set of pictures depicting objects, actions or concepts. Children are asked to name each picture by using a single word.

In addition, a more general assessment of oral language development or intelligence includes vocabulary tests. In English, for example, the Test of Word Knowledge (TOWK) (Wiig & Secord, 1992) is a clinical language assessment that identifies children with semantic weakness and low vocabulary, including receptive and expressive measures. The Picture Vocabulary of the Woodcock Johnson-III (WJ-III), in English (Woodcock et al., 2001); WJ-IV (Schrank et al., 2014) or the Kaufman Assessment Battery for Children (K-ABC), in English (Kaufman et al., 2018); in

German (Kaufman et al., 2014); in French (Kaufman & Kaufman, 2008) are also used in clinical practice to assess the intellectual and cognitive abilities and academic achievements of children (2 years and up). This assessment includes vocabulary measures similar to the PPVT and the EVT. In French, l'Epreuve de compréhension de Carrow-Woolfok (Carrow-Woolfolk, 1995) is the French-Quebec translation of the Test for Auditory Comprehension of Language (TACL) (Carrow-Woolfolk, 1985) and assess receptive skills for three forms of language (e.g., vocabulary, grammar and syntax) including a receptive measure. The task asks children to point out the picture that represents the meaning of words or sentences.

4.2 *General measures of depth of vocabulary*

While one-word picture identification or naming tasks are used to estimate the number of known words (e.g., breadth of vocabulary), tasks to examine the extent of word knowledge are used to reflect the gradual refinement of vocabulary knowledge. In contrast to breadth of vocabulary, depth is more difficult to operationalize since it is multidimensional (Nagy & Scott, 2000). For this reason, the extent of semantic representation is assessed by different tests covering all these dimensions.

The subtest Expressive Vocabulary of Wechsler Intelligence Scale for Children ((WISC-III) (in English (Wechsler, 1991); WISC-IV (Wechsler, 2003); adapted in e.g., in French, in Spanish (Wechsler, 2005a; Wechsler, 2005b)) or the subtest of Expressive Vocabulary: Word definitions in the Test Of Word Knowledge (TOWK) (in English (Wiig & Secord, 1992)) are commonly used. In these tasks, children were asked to define target words. Data scoring is based on the number of semantic features according to the semantic category of words. The Wechsler Abbreviated Scales of Intelligence (WASI) (in English (Wechsler, 2011)) and the Clinical Evaluation of Language Fundamentals (CELF), Fifth Edition (CELF-5) (in English (Wiig et al., 2013); CELF-4 in Dutch, (Kort et al., 2008); in French-Canadian (Wiig et al., 2009)) are also used and are similar to the WISC-III or the TOWK. The CELF includes subtests to obtain a fundamental language score in which children are asked to provide definitions (13-16 years) or indicate connections between words related by semantic features (4-16 years). This test is intended to be used to identify, detect and monitor language and communication problems. Alternately, the Multiple Meanings subtest of the Language Processing Test-Revised (LPT-R) (in English (Richard & Hanner, 2005)) asks children to define words in different contexts. This test evaluates children's ability to use synonyms or definitions. The TOWK (Wiig & Secord, 1992) also has a similar subtest in which participants have to provide two meanings for a target word. Finally, the Attributes subtest of the LPT requires children to describe attributes of names such as color, size, shape, or category (Richard & Hanner, 2005).

Researchers also use synonym tasks in addition to the definition task. For example, the receptive vocabulary task in the TOWK (in English (Wiig & Secord, 1992)) requires participants to choose the meaning of figurative expressions in a multiple-choice format or to select the synonym of a target word from four distractors. Similarly, in the Gates MacGinitie Vocabulary Subtest (MacGinitie et al., 1989) children have to choose the word or sentence that is closest to the meaning of the target word. As word knowledge is also related to word production, it seems interesting to evaluate fluency when assessing vocabulary. For instance, Tannenbaum et al. (2006) have used the Word Use Fluency subtest of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (in English (Good & Kaminski, 2002); in French (Dufour-Martel & Good, 2009)) to evaluate the depth of vocabulary by asking children to correctly use words in sentences. Finally, Ouellette and Shaw (2014) have assessed the depth of vocabulary by examining the lexico-semantic organization of the semantic system with the Word Classes subtest of the Clinical Evaluation of Language Fundamentals, Fifth Edition (CELF-5) (Wiig et al., 2013) in which children had to indicate among four words which were related by semantic features.

4.3 *Computer-based Tools*

Tools that assess the breadth of vocabulary are mainly in paper-and-pencil format. However, many researchers are stressing the advantages of using new technologies to access more precise and reliable behavioral measures (see Appendix A for a summary). Increasingly tests are computerized (e.g., PPVT-IV, EVT-3, DIBELS, CELF-5). In France, most computer-based assessments are part of child speech therapy assessment. The Bilan Informatisé de Langage Oral ((BILO) [Computerized checkup for oral language] (Khomsi et al., 2007a)) is a tool for assessing oral language among children (Khomsi et al., 2007a) and adolescents (Khomsi, et al., 2007b). BILO examines the development of oral

language skills before and after written language has been learned, in order to discover what holds children back in their development. These two modules assess the lexicon in terms of both reception and production with a procedure that is identical to the PPVT and the EVT.

The advantage of such a tool is that it provides complete results available in the form of comprehensive profiles. The Batterie Langage oral, Langage Ecrit, Memoire, Attention ((L2MA-2) [Oral language, written language, memory, attention] (Chevrie-Muller et al., 2011)) is also a French computer-based assessment used during speech assessments. The tool provides an assessment of oral and written language, skills associated with reading, memory and graphomotricity. These assessments include a lexicon module "*lexique en production*" [*lexicon in production*] in which children are asked to name pictures.

5. The advantage of new technologies for assessment

Few computer-based tools are currently available to assess vocabulary yet digital technologies have multiple advantages for assessment such as increased validity and reliability, increased engagement in the task by test-takers, easier implementation in the classroom, ability to record their strategies and performance measures and the development of new seemingly very promising adaptive assessment tools. The arguments in favor of digital technologies put forward here mostly come from research on general literacy assessment but they are also believed to be appropriate for specific vocabulary assessment (e.g., Cordero et al. (2018) for the use of touchscreen in writing and reading practices in children in Spanish). Moreover, arguments for digital technologies in literacy are mostly based on computer-based tools whereas touchscreen tablet applications appear to be equivalent to or even more advantageous than computer software.

5.1 *Increased validity and reliability*

Cognitive skills are not directly observable and measuring them is not easy since they depend on the task built by the researcher on the basis of a robust theory involving clear concepts. However, some criteria make it possible to check the efficacy and appropriateness of a psychological test, in particular with regard to validity and reliability. Validity indicates the extent to which the test measures the cognitive skills it was designed for (Smith, 2005).

The use of computer-based tools makes it possible to get multiple and precise measurements such as the number of correct answers and response times (Richter et al., 2013). Recording these scores is a necessary condition for measuring cognitive skills while the precision of these scores is a necessary condition for a valid assessment tool. Reliability refers to the ability of the task to reflect the real level of the cognitive skills assessed (Mowbray et al., 2003) based on, for example, the fidelity of task implementation. In enabling strict control of test conditions (duration of item presentation, automatic measure recording), using computer-based tools is a good way of increasing test reliability (Csapó et al., 2014).

5.2 *Increased engagement in the task by test-takers*

Evaluating cognitive processes requires engagement in the tasks by test-takers. Indeed, assessment results will vary depending on the attentional resources engaged. Due to their attractive aspect, based in part on the use of multimodal items (i.e., visual/audio items; see Cordero et al. (2018), Wang et al. (2008)), computer-based tools encourage test-taker's motivation and engagement in the task ((Chua & Don, 2013; Kucirkova et al., 2017) for different facets of engagement) compared to conventional paper-and-pencil tests (Singleton, 2001). Moreover, the interactive nature of computer-based tools provides sufficient guidance to test-takers. Therefore, greater autonomy (Protopapas & Skaloumbakas, 2007) results in an increased level of engagement.

5.3 *Easier implementation in the classroom*

Standardized presentation of items and greater test-taker engagement are two factors that facilitate implementation of computer-based assessment in the classroom. Moreover, scores are recorded and can be computed automatically (Schatz & Browndyke, 2002). This enables precise evaluation even in collective sessions. For this purpose, the

use of apps on touchscreen tablets is even more appreciable: Easily transportable (Falloon, 2013; Hutchison et al., 2012), familiar to children and highly intuitive (Beschorner & Hutchison, 2013), touchscreen tablets are particularly appropriate for young children (Couse & Chen, 2010; Walker, 2015) and ensure valid and reliable vocabulary measurement as well as the easy building of anonymized data banks for research or test standardization purposes (Schaefer et al., 2015).

5.4 *Recording learners' strategies*

In addition to increased reliability and easier classroom implementation, recording multiple and precise measures makes it possible to track test-taker's behavior (Jeong, 2014). Just by using a simultaneous interpretation of response times and correct answers, practitioners can distinguish different test-takers' strategies (Gil et al., 2015) or levels of motivation (Kong et al., 2007; Wise & Kong, 2005). Four interpretations can be made: (a) satisfactory processes with low response times and high correct response scores, (b) impaired processes with high response times and low correct response scores, (c) non-automated processes in the form of high response times and high correct response scores, or (d) impulsivity or "rapid guessing behavior" in the form of low response times and low correct response scores (Lee & Jia, 2014).

In reading comprehension literature, numerous measures are described as being informative of test-taker's strategies (reading times, number of text returns, for a more complete review of what can be done, see Greiff et al. (2016), Vidal-Abarca et al. (2011)). For example, in multiple-choice tasks, learner strategies and performance levels can be assessed depending on the way answer choices are displayed: Does the child wait to see all the possible answers before selecting one or do they select one as soon as they believe is the correct answer? When paired with response times and correct responses scores, it could inform on impulsivity or guessing behavior.

5.5 *Towards the adaptive assessment*

There is a debate on which kind of apps could be used (Kucirkova et al., 2017) or which signification has the use of the touchscreen for children (Rowsell & Wohlwend, 2016). Nevertheless, we can wonder how the apps could be adapted to individuals.

Adaptive testing refers to assessment tools that tailor item difficulties to subject ability levels (Tseng, 2016). The computer will propose different items depending on the success or failure in replying to the initial items. This kind of assessment enables the length of the assessment to be shorter (subjects do not see all items) and prevents the presentation of items that are excessively easy for high ability subjects as this could foster tiredness or excessively complex items for low ability subjects that could trigger loss of motivation and self-confidence.

Adaptive testing, therefore, enables more accurate evaluation than conventional testing and offers greater discrimination between test-taker's performance levels (Tseng et al., 2006). However, to be fully effective and offer the possibility of global scoring, adaptive testing requires an items bank to be built with items of increasing difficulty. With respect to this, Item Response Theory (IRT) (Baker, 2001) seems to be a good but complex way of building such an items bank ((da Silva Menegon et al. (2017), Makransky et al. (2017), Vloedgraven & Verhoeven (2009) who used IRT in various research fields). Contrary to conventional test theory that computes correct responses, the IRT test is based on estimating a subject's success or failure for each item that is weighted by the degree of difficulty. Therefore, an appropriate evaluation can be proposed to subjects by selecting appropriately difficult items and providing their estimated performances for those particular items. This enables their estimated performances for other item levels to be inferred.

6. Recommendation for vocabulary assessment

Reviews or studies on vocabulary assessment have identified the critical aspects to be taken into account when dealing with vocabulary assessments. Since vocabulary is multifaceted, testing vocabulary knowledge in-depth implies integrating multiple types of assessment into the evaluation (Kremmel & Schmitt, 2016). Vocabulary assessment procedures should closely match instructional goals (Coombe, 2011; Dougherty Stahl & Bravo, 2010) in order to

deliver more appropriate conclusions about the instruction needs or effectiveness (National Reading Panel, 2000; Watkins & DeThorne, 2000). Practitioners should be transparent to test-takers as to what the task is based on and what is expected of them (Coombe, 2011; National Reading Panel, 2000). Measures should be valid and reliable (Watkins & DeThorne, 2000). Result interpretation should take into account the fact that vocabulary knowledge is heavily dependent on life experience (Watkins & DeThorne, 2000). Test developers and practitioners should keep in mind that vocabulary knowledge is not only knowledge of the meaning-form relation but also essential to comprehension and general language ability (Oakhill et al., 2019). This implies proposing word knowledge assessments in embedded and comprehensive dimensions as well as linking vocabulary knowledge to passage comprehension assessment to measure the depth of vocabulary. Test developers should pay particular attention when generating items and distractors in the multiple-choice format. Practitioners should give great care to interpret and report results: The assessment's objectives and targeted vocabulary dimensions should be clearly indicated by test developers (Kremmel & Schmitt, 2016). To achieve this, links between vocabulary tests and other language or literacy tests should be given (Ouellette, 2006).

Vocabulary assessment allows us to know which vocabulary's facets (i.e., breadth and/or depth) would be effective or not in children. Therefore, depending on children's performances in the tests, training sessions to stimulate vocabulary could be proposed to improve it.

7. How to teach vocabulary?

7.1 *What research shows?*

7.1.1 *Vocabulary acquisition*

Increasing vocabulary does not only mean increasing the number of known words. It also involves increasing vocabulary depth and building a semantically organized network (Ordóñez et al., 2002). Acquiring word knowledge, therefore, requires building a stable mental representation with phonological, orthographic and semantic levels (Cartwright, 2007; Perfetti & Hart, 2002). Vocabulary can be acquired through three principal methods ranging from implicit to explicit learning: (a) incidental (implicit) learning, (b) embedded explicit instruction, or (c) extended explicit instruction (see Appendix B for a summary of the reviewed research in the current study).

7.1.2 *Incidental learning (implicit)*

Vocabulary knowledge is first learned incidentally based on the extraction of invariants (semantic, affective, perceptive, phonetic, and graphemic invariants) by abstracting elements common to linguistic occurrences (Christ & Chiu, 2018). This incidental learning requires that children receive multiple exposures (from 40 to 200 depending on the characteristics of the child, word types and learning context (Wasik et al., 2016)) in multiple contexts in order to go from recognizing the word to having a full understanding of the word (Biemiller, 2006). One way to favor incidental learning by multiplying word exposure is through storybook reading. For younger children (before they start to learn to read), storybook reading is shared (i.e., orally presented, read aloud by parents (Sénéchal et al., 2017)). For older children (able to read), storybook reading consists of wide (free) reading by children themselves.

The primary way for children to be exposed to new vocabulary words comes from the oral context of language experiences such as shared storybook reading (Lenhart et al., 2018). These instructions are mainly indirect and integrated into narrative contexts in which children have to listen to a story read to them. In this view, the contexts in which words are encountered contribute to a partial understanding of the meaning of the words. Many studies examine the effects of listening to stories on vocabulary growth (see Appendix B). These studies have shown that children can learn the meaning of unfamiliar words through incidental exposure during shared storybook reading (Biemiller & Boote, 2006), but this effect was limited. Researchers began to isolate factors to improve incidental vocabulary learning. In this context, studies have shown that repeated reading of stories (Biemiller & Boote, 2006; Penno et al., 2002) and rich discussions (Sénéchal, 1997; Wasik & Bond, 2001) led to an increase in vocabulary knowledge.

When children learn to read, their ability to derive the meaning of words from an oral context extends to the written context (Herman et al., 1987; Jenkins et al., 1984; Nagy et al., 1987). Children learn new words by reading on their own. Basically, the more children read, the more words they will meet, and the more they will be likely to learn. In some

studies, investigators have attempted to link children's reading volume to specific outcomes such as vocabulary. These studies have demonstrated that reading has a significant effect on vocabulary development (Cunningham & Stanovich, 1991). It is assumed that such a method increases the opportunity to encounter news words in different contexts.

7.1.3 *Toward direct (explicit) vocabulary instruction*

By calculating the averages from a range of studies, Biemiller (2006) postulates that, between the ages of 1 to 7, children acquire about 860 root word meanings per year. This means that at the end of grade 2, they know about 6,000 root word meanings. However, the average number of roots known stretches from 4,000 words for the 25 percent of children with the smallest number of known word meanings to 8,000 words for children in the highest quartile. According to Biemiller (2006), this difference is largely due to home support not the children's capacities or school support. Moreover, vocabulary learning depended on prior vocabulary knowledge: Children with low prior vocabulary knowledge are less receptive to vocabulary learning than those with high prior vocabulary knowledge (Webb & Chang, 2015). Then, the gap between children with low versus high home support will tend to expand if no targeted and differential intervention is proposed to children at risk (Coyne et al., 2004; Silverman, 2007; Spencer et al., 2012). So while word meanings seem to be learned incidentally based on vocabulary exposure, direct instruction methods are needed to fill the gap between children experiencing high and those experiencing low vocabulary exposure (Coyne et al., 2004; Justice et al., 2005; Penno et al., 2002; Robbins & Ehri, 1994; Sénéchal et al., 1995). Moreover, direct instruction of individual words should help children learn difficult words that are not part of daily experience. Studies in the field of vocabulary learning clearly identify two effective methods: the embedded and the extended methods.

Only a small number of studies have examined the effect of direct learning in the context of shared book reading (see Appendix B). The most common approach applied was to provide children with explanations of words when they were encountered in repeated story reading. The aim of this "*embedded*" approach that focuses on breadth vocabulary growth is to introduce a large number of words to children within the context of story read-aloud (Biemiller & Boote, 2006; Justice et al., 2005; Penno et al., 2002; Spencer et al., 2012).

According to a review of 50 experimental or quasi-experimental vocabulary studies (National Reading Panel, 2000), learning words in context results in better vocabulary acquisition than learning words in isolation. Hence, two types of strategies would theoretically be effective for vocabulary instruction. First, embedded strategies that consist in teaching the meaning of words encountered in context. Storybook reading is one of the most documented embedded strategies for preschool or primary children (for a recent review on storybook reading (Wasik et al., 2016)). However, variations can be found within storybook strategy depending on different conditions: The number of reading sessions, on the one hand, the presence or absence of explanations given by adults on the other (incidental versus explicit learning).

In a review of seven studies assessing the effectiveness of storybook reading on vocabulary acquisition based on these two conditions, Biemiller (2006) argued that (a) vocabulary gains were lower if storybooks are read only once (about 4% more words understood by children after compared to before story reading) compared to the storybook read three of four times (10% to 15% more words understood by children after compared to before story reading). He also argues that (b) vocabulary gains were higher when adult explanations of new word meanings were taken into account (14% to 29% more words understood by children after compared to before story reading). Following this review, Biemiller and Boote (2006) conducted a study with 43 children from kindergarten to Grade 3. In this study, the authors examined the effect of repeated reading (twice versus 4 times) with the presence or absence of explanations on new vocabulary acquisition. Vocabulary knowledge before and after the test was assessed by asking children to provide explanations of words from context sentences. After the test, researchers found that an average gain of 12% was obtained when stories were repeated. When brief explanations were provided, a further 10% gain was added. Overall, children learned more words when direct instructions were provided (22%) relative to incidental instruction (12%). In a second study, Biemiller and Boote (2006) show that more words were learned (41% in study 2 and 22% in study 1) by adding two reviews of word meanings. This new presentation provided children with the opportunity to refine the meaning of words in new contexts not based on the book and with additional explanations of word meanings.

7.1.4 *Extended vocabulary instruction*

Extended strategies consist in teaching the meanings of words encountered in context then extending the teaching

of these words in other contexts, teaching collocations, synonyms, antonyms, or polysemies. For example, McKeown and their colleagues (2014; 1983) conducted studies that investigated the effect of extended vocabulary instruction on reading comprehension and word knowledge in fourth graders. The instructional program includes activities that require children to (a) interact with words and their meanings in order to use them, (b) associate words with their contexts, (c) generate contexts for target words and (d) compare words in order to examine their relationships.

The assessment measures focused on accuracy (children were instructed to choose the correct definition of target words among four alternatives), fluency (children were asked to decide if a target word belonged to a semantic category) and reading comprehension (children have to read then recall and answer questions about the stories). The results of this study have shown that the group of instructed children showed a significant advantage in terms of both vocabulary knowledge and comprehension compared to the group of non-instructed children.

The review of the National Reading Panel (2000) also added that incrementally teaching a word meaning was useful in developing depth dimensions in addition to vocabulary breadth and that heterogeneity of context in which words are encountered was essential to "*precise, nuanced and even sophisticated*" (p. 290) meanings of newly acquired words. Hence, extended strategies seem more appropriate than embedded strategies to teach high-quality vocabulary knowledge.

This assumption is consistent with the results of the Silverman (2007) study who compared the effectiveness of three instruction types: (a) contextual instruction that consists in teaching words in context and relating them to the children's personal experience, (b) analytical instruction that consists in enhancing contextual instruction (adding semantic analysis of words in other contexts and children's experience), and (c) anchored instruction that consists in explicitly linking words to their phonological and orthographic properties. Contextual instruction could be assimilated to an embedded strategy, whereas analytical and anchored instructions that add more contextual use could be assimilated to extended strategies.

Silverman (2007) compared results in vocabulary tests in three groups of children in kindergarten (two classrooms per group) experiencing the three different approaches using a pre-test/training/post-test approach. The results showed that performance in learning targeted vocabulary was higher in analytic and anchored instruction (assimilated to extended strategies) compared to contextual instruction (assimilated to an embedded strategy). These results are consistent with those of Coyne et al. (2009) who conclude that embedded strategies provided partial knowledge of a word (e.g., the meaning-form link that could be assimilated to vocabulary breadth), while extended strategies provided a finer and more precise knowledge of a word (that could be assimilated to vocabulary depth).

Recently, many researchers have used active processing to promote vocabulary instruction in young children (Beck & McKeown, 2007; Coyne et al., 2009). These studies have reported a positive impact of extended vocabulary instruction on word learning, but the effects on reading comprehension have not been examined or were not significant (McKeown & Beck, 2014). Only Coyne and colleagues (2010) have reported a positive effect on listening comprehension measured for children who learn the meaning of words through deeper processing.

Finally, active processing also refers to activities in which children examine how words are related to each other through semantic mapping (Johnson et al., 1982) and semantic feature analysis (Bos et al., 1989). In the first method, children were instructed to draw a graph centered on a given concept the connections of which are ideas that explain the concept. While building these maps, discussions between children and teachers are fundamental to successful construction. Another method used in learning words is semantic feature analysis. Children are asked to draw a summary table where category members would be listed in the first left column and the characteristics of these members in the first row of the table. The children must indicate whether the members have the characteristics (a "+" is drawn in the corresponding box) or not (a "-" is drawn in the corresponding box). The aim of such a method is to help students categorize vocabulary words and distinguish similarities and differences between related ideas.

In short, extended vocabulary instruction relies on three strategies: multiplying word exposure in various contexts, learning words incrementally, and using active learning. One limit of extended strategies is the fact that, when compared to embedded strategies, extended strategies are time-consuming (Coyne et al., 2009), therefore, should be limited to teaching targeted words and vocabulary learning for children at risk due to their low initial vocabulary levels. This, therefore, implies implementing a differentiated pedagogy based on initial levels and this requires an assessment of initial levels. In one of his studies, Coyne et al. (2010) examined the effectiveness of direct extended instruction for children considered as being at risk. Three conditions were set as being necessary for extended instruction: new

word meanings are learned through simple definitions and synonyms, intervention provides extended opportunities to encounter the word in various contexts, therefore, promoting depth of vocabulary. Children were separated into two groups: The experimental one that received 18 weeks of direct extended instruction (36 hours) targeting 54 words to be learned and the control group that did not receive the instruction. Results showed that children in the experimental groups obtained better results than those in the control groups over several measures (generalized receptive vocabulary and listening comprehension). However, the effectiveness of the intervention was closely related to initial receptive vocabulary: Children with low initial receptive vocabulary benefit less from the intervention than those with high initial receptive vocabulary. This finding illustrates a Matthew effect in vocabulary learning that reinforces the need for early intensive vocabulary intervention for children at risk.

7.1.5 *What to teach?*

Vocabulary learning is not only recognizing new words but rather implies being able to understand and use them appropriately and fluently in context (Ordóñez et al., 2002). Also, teaching vocabulary requires teaching several dimensions (Bromley, 2004). The choice of intervention and the dimensions taught will depend on the learning goals. As presented in the study of Coyne et al. (2009), vocabulary breadth interventions should be limited to teaching word general knowledge whereas vocabulary depth interventions should be limited to teaching targeted vocabulary. The question is then to define which targeted words are needed to be taught with extended interventions?

7.1.6 *Which words to teach?*

Selecting words to teach is one major concern when designing new vocabulary teaching systems. Should the intervention focus on unknown words? Should it focus on improving knowledge of frequent words or encourage exposure and learning of rare words? Or should the intervention focus on important and useful words only? In the latter case, what criteria are applied when choosing the words considered as useful and important? With respect to this, Hiebert (2005) described the characteristics of an "*efficient and effective component*" of vocabulary. Efficient takes account of the wealth (number of semantic associations, word frequency and morphological families), effective takes account of the probability of encountering these words depending on the grade level, and component refers to the fact that selected words only constitute a part of the vocabulary that is required at a targeted grade level.

Different methods are described to select words to teach (Gray & Yang, 2015). The most current one is the Tier-2 method described by Beck et al. (2005). A corpus of words in a text are divided into tiers: Tier-1 comprises highly frequent words that are probably already known by students, Tier-2 comprises words needed to be used by subjects for mature language (i.e., subject knows a concept and the Tier-2 word can help them express simply their idea using appropriate words associated with this concept) and Tier-3 comprises rare word linked with specific domains. According to the authors, vocabulary intervention should focus on Tier-2 words in order to facilitate comprehension or expression.

Biemiller (2005) proposed another approach that consists of focusing intervention on words known by 40 to 80% of students at a targeted grade level (end of grade 2). These words are then considered at an intermediate level and teaching them would help reduce the gap between low and high vocabulary learners. In a recent review, Gray and Yang (2015) compared the advantages and disadvantages of five approaches to select words to be taught including Tier-2 (Beck et al., 2005) and Biemiller's (2005) approaches. The principal advantage of the Tier-2 approach lies in the flexibility of the approach (words can be divided into tiers for each text studied in class) while the main drawback is the lack of a common core of words known to all children of the same age or grade-level. In Biemiller's approach, the advantages and disadvantages are reversed: the words to be taught are the same for all children but this is performed independently of the material used to enhance vocabulary learning. The difficulty, therefore, lies in implementing words to be taught into the school sequence.

7.1.7 *Teaching word meanings and/or learning word meanings*

Another way to question what to teach in vocabulary instruction examines the more general problem of teaching a specific knowledge or teaching meta-abilities such as strategies to learn new vocabulary. Vocabulary intervention studies are critically missing tests of learning strategies and measures of transfer knowledge in general language and literacy

(Coyne et al., 2010). Therefore, it is important to differentiate between (a) interventions that target the acquisition of specific word meanings, (b) interventions that focus on teaching more general word knowledge, and (c) interventions providing instruction in learning strategies.

In a dated study for second language learners, Schmitt and Schmitt (1993) suggested that students need to have the skills to help manage their own learning. They identify various strategies for learning new words such as "using reference materials" (e.g., dictionary), "asking others for information" (e.g., the teacher, classmates), "analyzing words from available information" (e.g., inferring the meaning of a word from its context), or "creating a system to analyze words". Then, they present strategies to remember the meanings of words learned, such as "repetition" or "studying the formal and grammatical aspects of a word". In this study, Schmitt and Schmitt (1993) stressed the fact that low vocabulary learners were less likely to use multiple and various strategies to learn new word meanings than high vocabulary learners. Moreover, the use of vocabulary strategies seems to have an impact on more general language ability and favors autonomy in vocabulary learning. These results confirm the need to teach vocabulary learning strategies to low vocabulary students in order to foster the use of these strategies. New studies of this type would be invaluable in assessing the impact of teaching vocabulary learning strategies to young children in a first language learning context.

8. Advantage of new technologies in teaching vocabulary

8.1 *Increased motivation*

The first arguments in support of the use of digital technologies in learning concern motivation. Indeed, computers or tablets are attractive to children due to their fun aspect (e.g., games). Moreover, the multimodal items and interactivity (Stetter & Hughes, 2011) offered by these technologies encourage high attentional engagement by learners (Beschorner & Hutchison, 2013; Sung et al., 2008). Finally, computer-based or tablet-based technologies have benefits for children with learning difficulties or disabilities (Fernández-López et al., 2013; Stetter & Hughes, 2011). Students with learning disabilities or atypical profiles often show poor results despite their efforts. As a result, these children have less self-confidence and motivation to learn and to go to work in school. The appealing aspect of digital technologies makes it possible to (re)motivate these students. Indeed, recently, Mize et al. (2018) have shown that visual and auditory support (image, visual animations and graphics) have a positive effect on the learning outcome for children with learning disabilities. In addition, the fact that digital programs provide individualized training, respect individual learning speeds and give non-judgmental and systematic feedback, can enable students who have dropped out to regain self-confidence and re-engage in learning processes (Lynch et al., 2000).

8.2 *Favoring processing acquisition: the advantages of controlled feedback*

The presence of feedback in a digital tool dedicated to learning is one of the conditions for its efficiency and superiority compared to a paper-and-pencil tool (Babin et al., 2009; Johnson-Glenberg, 2007; Lynch et al., 2000; Sung et al., 2008; Van der Kleij et al., 2012). Indeed, adding systematic feedback following the learner's answers provides for interactive (the subject receives feedback on their answers during the task) and personalized (the feedback is dependent on the subject's answers) learning (Van der Kleij et al., 2012). Feedback differs based on four complementary aspects: the types of feedback when the feedback is provided and the length and complexity on the one hand (Shute, 2008), and the level of feedback impact on the other (Hattie & Timperley, 2007).

Three main types of feedback are defined: Firstly, "knowing results" feedback, in which the feedback provided by the software simply indicates whether the answer is right or wrong without giving the correct answer. The second type of feedback is corrective ("knowing correct response"). The third type corresponds to elaborative feedback: The feedback gives information to succeed in the exercise and thus implies a meta-cognitive component favoring the spreading of learning to other contexts. The feedback can also be defined by the timing and this is on two levels: Immediate, the feedback is provided directly after the question or is delayed, the feedback is provided at the end of the exercise for example. The complexity of the feedback is part of a continuum from simple to complex and takes into account, in particular, the type and the length of the feedback.

Four points of impact are targeted by the feedback (Hattie & Timperley, 2007): they can be related to the software user ("self"), the task ("task"), the general procedure necessary to succeed in the task ("processing") or the strategy chosen by the subject ("regulation"). The choice of feedback impact level may slightly change the content of the feedback without changing the type. For example, elaborate feedback will vary depending on whether it is "task" or "procedure" oriented: In the first case, the help given will be relative to the task such as giving contextualized clues (e.g., keywords to answer an understanding question); in the second case, the help given will concern the procedure such as giving the general rule for the success of this type of exercise. Result feedback may relate to the task (e.g., "the answer is wrong") or the user with congratulations or encouragement (e.g. "You made a mistake. Take back your chance!").

The effectiveness of the feedback depends on the interaction between the characteristics of the user (age, developmental level and attitude), the characteristics of the task and the characteristics of the feedback defined in the previous paragraph. Thus, there is no one type of feedback superior to all others. Nevertheless, Stobart (2008) proposes three general criteria conditioning the effectiveness of feedback for learning: first, the learner must need feedback. For example, Hattie and Timperley (2007) discuss the need for feedback when there is too much difference between the goal and the understanding of that goal. Secondly, the learner must have the time to receive the feedback: Feedback that disappears too quickly from the screen will not be effective. Third, the learner must want to pay attention to feedback and be able to use it. Nor will excessively complex feedback be effective. According to Timmers and Veldkamp (2011), users pay more attention to feedback based on incorrect responses compared to feedback following a correct answer. The effectiveness of learning through feedback thus seems to be partly based on the fact that the user makes mistakes.

In a more recent study, Van der Kleij et al. (2015) compared the effectiveness of different feedback depending on types (i.e., elaborated, knowing response, or knowing correct response), level of learning outcomes (i.e., high or low), timing (i.e., immediate or delayed) and subjects age (i.e., primary, secondary school or university). Results showed that elaborated feedback was more effective than knowing response and knowing correct response feedback, particularly for higher-order learning outcomes. They also demonstrated that feedback provided immediately was more effective than delayed feedback and that feedback was more effective for students from university than for younger children from primary and secondary school.

To conclude, the appropriateness of the feedback provided during an intervention is critical for its effectiveness. Digital technologies offer the opportunity to program targeted feedback adapted to targeted situations. Then, the feedback type and conditions must be consistent and systematic in order to respect learners' needs and ensure intervention efficacy.

8.3 *Allowing differentiated instruction*

One-to-one tutoring is considered by some authors as the most effective instruction form (Van der Kleij et al., 2012). Indeed, one-to-one tutoring makes it possible to actively process learning strategies, receive systematic, immediate and appropriate feedback and progress at the learner's own pace. Digital technologies have these three characteristics (McNamara et al., 2004; Sung et al., 2008). Children can therefore benefit from an individualized program even in a group situation at school (Johnson-Glenberg, 2007). This is consistent with the aim to offer a differentiated pedagogy in order to provide each child with instruction adapted to his/her skills and his/her initial level. Thus, the subject is an active agent in his/her learning (McNamara, 2004; McNamara et al., 2004; Rutten et al., 2012).

In addition, digital technologies offer the possibility of programming success conditions (objective achieved, skills acquired) to move on to the next exercise or to start an exercise over when the competence is considered as "not acquired" or "in the process of being acquired ". This reinforces the ability of digital tools to provide individually targeted instruction that could be efficient even in full classroom situations.

8.4 *With touchscreen tablets particularly*

Regarding the use of the touchpad, studies have highlighted ergonomic arguments such as mobility (touchscreen tablets are no bigger or heavier than a book), the ability to easily customize the interface with for example the ability to zoom, incorporate images or definitions, or highlight words as they are read (Hutchison et al., 2012). The intuitive nature of touchpad navigation enables digital tools to be used with younger children and eliminates the need for specific

teaching in how to use the tool (Neumann & Neumann, 2014).

Conversely, the difficulty added by the mouse and keyboard as intermediates between the hand and the screen limits the possibility of using computers with young children without prior awareness-raising and learning (Couse & Chen, 2010). Several authors define the characteristics that applications must possess in order to be effective. For example, Fernández-López et al. (2013) emphasize the notions of usability, accessibility, flexibility, adaptability and mobility. On the cognitive level, the main arguments put forward are the possibility of presenting multi-modal stimuli such as visual, auditory, tactile and kinesthetic stimuli (Chen & Sun, 2012). This favors the possibility of displaying multiple encoding which is acknowledged as being more robust, the possibility of compensating one deficit channel by another procedure (Fernández-López et al., 2013), the possibility of facilitating the understanding of instructions or navigation (e.g., navigation icons, instructions read aloud by the application).

9. Recommendations for (explicit) vocabulary instruction

Recommendations for vocabulary instruction can be divided into two parts: First, research has defined the conditions that favor vocabulary acquisition. Secondly, research has defined the conditions that favor the effectiveness of vocabulary instruction methods in enhancing the feasibility of active participation by vocabulary learners (Bromley, 2004).

For vocabulary acquisition, five main principles can be set down: (a) incidental learning enables more new word meanings to be learned than explicit teaching (Manyak et al., 2014), (b) new word meanings are learned more easily when presented in context (Silverman, 2007; National Reading Panel, 2000) and (c) when acquired incrementally (Nagy & Scott, 2000). This implies that (d) vocabulary learners should benefit from multiple exposures to targeted words and in various contexts and with various approaches in order to build in-depth vocabulary knowledge (National Reading Panel, 2000; Silverman, 2007). Finally, (e) instruction should target the acquisition of new word meanings and the teaching of learning strategies (Manyak et al., 2014).

As regards fostering the effectiveness of vocabulary instruction, Manyak et al. (2014) describe four principles. (1) Teachers should establish routines for introducing new word meanings. This makes it possible to reduce time spent on exercise because after the first session, additional time is no longer necessary to explain instructions. This also enables more autonomy to be provided to children in their activities, thereby enhancing children's engagement. However, in order to prevent disengagement, it is recommended to switch between several routines over the year. (2) Teachers should promote activities with deep processing of word meanings, (3) directly address confusion among learners and foster participation by each learner. For this latter purpose, the use of new technologies is advantageous since it makes it possible to provide individual, immediate, systematic and appropriate feedback to learners' answers as well as adaptive differentiated intervention.

10. Conclusion

The concept of vocabulary knowledge is complex. Depending on the researcher's point of view, vocabulary does not refer to the same construct. Yet a consensus seems to be taking shape on the issue of how to define vocabulary. Vocabulary can be investigated in its breadth and depth. Generally, to assess the number of known words, standardized tests examine the connection between the form and meaning of words in a receptive and expressive way. Because these tests are composed of an average of 100 items, it gives a general indicator of the size of the child's lexicon and identifies children at risk. But, breadth of vocabulary is not the critical dimension since the depth of vocabulary is the dimension that has been reported as fundamental for reading comprehension. The major problem with depth of vocabulary is that it reflects the entire complexity and multidimensionality of vocabulary knowledge as reported by Nagy and Scott (2000) for example. In other words, the quality of semantic knowledge about words cannot be limited to a simple test in which children are asked to define words. Generally, researchers used other tests to capture the multidimensional nature of the meaning of words such as synonym selection or word production, in line with a different theoretical approach.

A review of the scientific literature suggests that vocabulary development is a multidimensional process that

requires a combination of direct and indirect instruction, discussion and active learning processes. Given the large number of words that are learned during childhood and adolescence (Biemiller & Slonim, 2001; Nation, 2006), it can be argued that direct vocabulary teaching only explains a few of these words. Most of them are learned incidentally. This assumption is supported by experimental studies showing that word knowledge can be acquired through the written context (reading) and oral context (shared reading). However, some authors have sought to improve this learning by adding, on the one hand, repeated exposures to words as well as explanations of the meaning of words (embedded). These explanations involve brief definitions of word meanings. On the other hand, authors have also studied more interactive activities to improve vocabulary (extended). These methods are apparently effective in increasing vocabulary skills and, with respect to the latter, providing the components of a vocabulary learning method that is effective for both vocabulary and comprehension (National Reading Panel, 2000; Marulis & Neuman, 2013).

11. Perspectives

In line with the literature, several arguments must be taken into account when designing new high-quality vocabulary tools for assessment or instruction using digital technologies. In the context of vocabulary assessment, digital tools could provide accurate data collection and analysis of children's responses as well as response time and correct responses, an increased validity and reliability of assessment, and attractive support to enhance children's engagement in the task. These features are germane to the development of an accurate and detailed profile of children both in terms of vocabulary and reading skills. Moreover, future assessment tools should target adaptive assessment based on, for example, item response theory.

In the context of vocabulary instruction, computer-based tools should provide an interactive and multi-modal environment to favor motivation and engagement, systematic feedback appropriate to learners' characteristics such as age or initial vocabulary level to favor autonomy and active learning, and differentiated instruction depending on learners' initial vocabulary level enabling children to progress at their own pace without group or teacher pressure. More generally, the effectiveness of digital tools on learning requires more in-depth collaboration between researchers (cognitive aspects), teachers (pedagogical aspects) and developers (ergonomic aspects) to promote wider use of their potentialities and benefits.

Conflict of interest statement

The authors declare no conflict of interest.

References

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary development: A Morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10). https://doi.org/10.2307/1166112

Babin, L.-M., Tricot, A., & Mariné, C. (2009). Seeking and providing assistance while learning to use information systems. *Computers & Education, 53*(4), 1029-1039. http://doi.org/10.1016/j.compedu.2009.01.015

Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Publications.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251-271. http://doi.org/10.1086/511706

Beck, I. L., McKeown, M. G., & Kucan, L. (2005). Choosing words to teach. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and Learning Vocabulary: Bringing Research to Practice* (pp. 209-222). Routledge.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506-521. http://doi.org/10.1037/0022-0663.74.4.506

Beschorner, B., & Hutchison, A. (2013). IPads as a literacy teaching tool in early childhood. *Online Submission, 1*(1), 16-24.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary

grade vocabulary instruction. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223-242). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D.K. Dickinson, & Neuman, S.B. (Eds.), *Handbook of Early Literacy Research* (pp. 41-51), Guilford Press.

Biemiller, A., & Boote, C. (2006). An Effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44-62. http://doi.org/10.1037/0022-0663.98.1.44

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498-520. http://doi.org/10.1037/0022-0663.93.3.498

Bos, C. S, Anders, P. L., Filip, D., & Jaffe, L. (1989). The Effects of an interactive instructional strategy for enhancing reading comprehension and content area learning for students with learning disabilities. *Learning Disabilities, 22*(6), 384-390. http://doi.org/10.1177/002221948902200611

Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., Shankweiler, D. P. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing, 29*(3), 435-451. https://doi.org/10.1007/s11145-015-9608-6

Brinchmann, E. I., Hjetland, H. N., & Lyster, S.-A. H. (2015). Lexical quality matters: Effects of word knowledge instruction on the language and literacy skills of third- and fourth-grade poor readers. *Reading Research Quarterly, 51*(2), 165-180. https://doi.org/10.1002/rrq.128

Bromley, K. (2004). Rethinking vocabulary instruction. *Language and Literacy Spectrum, 14*, 3-12.

Brownell, R. (2001). *Expressive One-Word Picture Vocabulary Test-Spanish-Bilingual Edition*. Novato, CA: Academic Therapy Publications.

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique, 114*(4), 647-662. https://doi.org/10.4074/S0003503314004035

Carrow-Woolfolk, E. (1985). *TACL-R: test for auditory comprehension of language*. DLM Teaching Resources.

Carrow-Woolfolk, E. (1995). Oral and written language scales. *Circle Pines, MN: American Guidance Service, 93*(94), 947-964.

Cartwright, K. B. (2007). The contribution of graphophonological-semantic flexibility to reading comprehension in college students: Implications for a less simple view of reading. *Journal of Literacy Research, 39*(2), 173-193. http://doi.org/10.1080/10862960701331902

Chen, C.-M., & Sun, Y.-C. (2012). Assessing the effects of different multimedia materials on emotions and learning performance for visual and verbal style learners. *Computers & Education, 59*(4), 1273-1285. http://doi.org/10.1016/j.compedu.2012.05.006

Chevrie-Muller, C., Maillart, C., & Simon, A. M. (2011). L2MA-2: Langage oral, langage écrit, mémoire, attention [Oral language, written language, memory, attention]. ECPA.

Chiu, Y.-D. (2018). The simple view of reading across development: the prediction of grade 3 reading comprehension by prekindergarten skills. *Remedial and Special Education, 39*(5). https://doi.org/10.1177/0741932518762055

Christ, T., & Chiu, M. M. (2018). Hearing words, learning words: How different presentations of novel vocabulary words affect children's incidental learning. *Early Education and Development, 29*(6), 831-851. http://doi.org/10.1080/10409289.2018.1484648

Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior, 29*(5), 1889-1895. https://doi.org/10.1016/j.chb.2013.03.008

Colenbrander, D., Kohnen, S., Smith-Lock, K., & Nickels, L. (2016). Individual differences in the vocabulary skills of children with poor reading comprehension. *Learning and Individual Differences, 50*, 210-220. https://doi.org/10.1016/j.lindif.2016.07.021

Coombe, C. (2011). Assessing vocabulary in the language classroom. In D. Anderson & R. Sheehan (Eds.), *Focus on Vocabulary: Emerging Theory and Practice for Adult Arab Learners* (pp. 111-124). United Arab Emirates: HCT Press

Cordero, K., Nussbaum, M., Ibaseta, V., Otaíza, M. J., & Chiuminatto, P. (2018). Read, write, touch: Co-construction and multiliteracies in a third-grade digital writing exercise. *Journal of Computer Assisted Learning, 34*(2), 162-173. http://doi.org/10.1111/jcal.12224

Couse, L. J., & Chen, D. W. (2010). A tablet computer for young children? Exploring its viability for early childhood education. *Journal of Research on Technology in Education, 43*(1), 75-96. http://doi.org/10.1080/15391523.2010.10782562

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr, R., & Kapp, S. (2009). Direct vocabulary instruction in

kindergarten: Teaching for breadth versus depth. *The Elementary School Journal, 110*(1), 1-18. http://doi.org/10.1086/598840

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr, R., Ruby, M., Crevecoeur, Y. C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness, 3*(2), 93-120. https://doi.org/10.1080/19345741003592410

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145-162. https://doi.org/10.1207/s15327035ex1203_3

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology, 106*(3), 639. https://doi.org/10.1037/a0035756

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*(2), 264-274. https://doi.org/10.1037/0022-0663.83.2.264

da Silva Menegon, L., Vincenzi, S. L., de Andrade, D. F., Barbetta, P. A., Merino, E. A. D., & Vink, P. (2017). Design and validation of an aircraft seat comfort scale using item response theory. *Applied Ergonomics, 62*, 216-226. https://doi.org/10.1016/j.apergo.2017.03.005

Dale, H. C. A. (1965). Recency in two-choice guessing tasks. *British Journal of Psychology, 56*(1), 33-44. https://doi.org/10.1111/j.2044-8295.1965.tb00941.x

Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press Cambridge.

Dougherty Stahl, K. A., & Bravo, M. A. (2010). Contemporary classroom vocabulary assessment for content areas. *The Reading Teacher, 63*(7), 566-578. https://doi.org/10.1598/RT.63.7.4

Dufour-Martel, C., & Good, R. H. (2009). Investigating the psychometric properties of three French language early reading measures. *Effective Education, 1*(1), 87-100. https://doi.org/10.1080/19415530903043698

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Pearson Assessments.

Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *The British Picture Vocabulary Scale*. GL Assessment Limited.

Dunn, L. M., Dunn, L. M., & Theriault-Whalen, C. (1993). *EVIP: Échelle de vocabulaire en images Peabody* [*Pictures Vocabulary Scale Peabody*]. https://books.google.fr/books?id=WW_6oAEACAAJ

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imagenes Peabody: TVIP: Adaptacion Hispano americana* [*Peabody Picture Vocabulary Test: PPVT: Hispanic-American Adaptation*]. Circle Pines, MN: American Guidance Service (AGS), Inc.

Elley, W. (1989). Vocabulary acquisition from listening to stories. *ReadingResearch Quarterly, 24*(2), 174-187. https://doi.org/10.2307/747863

Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology, 115*(3), 497-516. https://doi.org/10.1016/j.jecp.2012.12.001

Falloon, G. (2013). Young students using iPads: App design and content influences on their learning pathways. *Computers & Education, 68*, 505-521. https://doi.org/10.1016/j.compedu.2013.06.006

Fernández-López, Á., Rodríguez-Fórtiz, M. J., Rodríguez-Almendros, M. L., & Martínez-Segura, M. J. (2013). Mobile learning technology based on iOS devices to support students with special education needs. *Computers & Education, 61*, 77-90. https://doi.org/10.1016/j.compedu.2012.09.014

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics, 28*(2), 241-265. https://doi.org/10.1093/applin/amm010

Gil, L., Martinez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers & Education, 82*, 50-59. https://doi.org/10.1016/j.compedu.2014.10.026

Good, R. H., Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills: DIBELS*. R. H. Good (Ed.). Dynamic Measurement Group.

Gray, S., & Yang, H.-C. (2015). Selecting vocabulary words to teach. *SIG 1 Perspectives on Language Learning and Education, 22*(4), 123-130. http://doi.org/10.1044/lle22.4.123

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior, 61*, 36-46. https://doi.org/10.1016/j.chb.2016.02.095

Groupe coopératif en orthophonie-Région Laval, Laurentides, Lanaudière. (1995). *Test de dénomination EO-WPVT-R de Gardner* [*Expressive One-Word Picture Vocabulary Test*]. Ordre des orthophonistes et audiologistes du Québec

[College of Speech Language and Audiology of Quebec], Montréal, Québec, Canada.
Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL-International Journal of Applied Linguistics, 166*(2), 278-306. https://doi.org/10.1075/itl.166.2.04gyl
Harsch, C., & Hartig, J. (2016). Comparing C-tests and Yes/No vocabulary size tests as predictors of receptive language skills. *Language Testing, 33*(4), 555-575. https://doi.org/10.1177/0265532215594642
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487
Herman, P. A., Anderson, R. C., Pearson, P. D., & Nagy, W. E. (1987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly, 22*(3), 263. https://doi.org/10.2307/747968
Hiebert, E. H. (2005). In pursuit of an effective, efficient vocabulary curriculum for elementary students. In E.H. Hiebert & M. L. Kamil (Eds.). *Teaching and Learning Vocabulary: Bringing Research to Practice* (pp. 243-264). Routledge.
Hoffman, J. L., Teale, W. H., & Paciga, K. A. (2014). Assessing vocabulary learning in early childhood. *Journal of Early Childhood Literacy, 14*(4), 459-481. https://doi.org/10.1177/1468798413501184
Hutchison, A., Beschorner, B., & Schmidt-Crawford, D. (2012). Exploring the use of the iPad for literacy learning. *The Reading Teacher, 66*(1), 15-23. https://doi.org/10.1002/TRTR.01090
Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*(4), 767-787. https://doi.org/10.3102/00028312021004767
Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology, 33*(4), 410-422. https://doi.org/10.1080/0144929X.2012.710647
Johnson, D. D., Toms-Bronowski, S., & Pittelman, S. D. (1982). An investigation of the effectiveness of semantic mapping and semantic feature analysis with intermediate grade children (Program Report). Madison: University of Wisconsin, Center for Educational Research.
Johnson-Glenberg, M. C. (2007). Web-based reading comprehension instruction: Three studies of 3D-readers. In D. S. McNamara (Eds.), *Reading Comprehension Strategies: Theories, Interventions, and Technologies* (pp. 293-324). Routledge.
Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An Efficacy study with at-risk kindergartners. *Language Speech and Hearing Services in Schools, 36*(1), 17. https://doi.org/10.1044/0161-1461(2005/003)
Kaufman, A. S., & Kaufman, N. L. (2008). *KABC-II batterie pour l'examen psychologique de l'enfant, deuxième ed*. [*Kaufman Assessment Battery for Children-Second Edition*]. ECPA.
Kaufman, A. S., Kaufman, N. L., Drozdick, L. W., & Morrison, J. (2018). *Kaufman Assessment Battery for Children-Second Edition Normative Update (KABC-II NU), manual supplement*. Bloomington, MN: NCS Pearson.
Kaufman, A. S., Kaufman, N. L., Melchers, P., & Melchers, M. (2014). *German-language adaptation of the Kaufman Assessment Battery for Children-Second Edition*. Frankfurt, Germany: Pearson.
Khomsi, A., Khomsi, J., & Pasquet, F. (2007a). *Bilan Informatisé de Langage Oral au cycle 2 (BILO-2)* [*Computerized assessment of oral language for 5-7 years old children*]. Paris: Editions du Centre de Psychologie Appliquée.
Khomsi, A., Khomsi, J., & Pasquet, F. (2007b). *Bilan Informatisé de Langage Oral au cycle 3 (BILO-3C)* [*Computerized assessment of oral language adolescents*]. Paris: Editions du Centre de Psychologie Appliquée.
Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing, 25*(2), 347-373. https://doi.org/10.1007/s11145-010-9272-9
Kong, X. J., Wise, S. L., & Bhola, D. S. (2007). Setting the response time threshold parameter to differentiate solution behavior from rapid-guessing behavior. *Educational and Psychological Measurement, 67*(4), 606-619. https://doi.org/10.1177/0013164406294779
Kort, W., Schittekatte, M., & Compaan, E. (2008). *CELF-4-NL: Clinical Evaluation of Language Fundamentals-Vierde-Editie*. Amsterdam: Pearson Assessment and Information.
Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly, 13*(4), 377-392. https://doi.org/10.1080/15434303.2016.1237516
Kucirkova, N., Littleton, K., & Cremin, T. (2017). Young children's reading for pleasure with digital books: Six key facets of engagement. *Cambridge Journal of Education, 47*(1), 67-84. https://doi.org/10.1080/0305764X.2015.1118441

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Lee, Y.-H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-Scale Assessments in Education, 2*(1), 8. https://doi.org/10.1186/s40536-014-0008-1

Lenhart, J., Lenhard, W., Vaahtoranta, E., & Suggate, S. (2018). Incidental vocabulary acquisition from listening to stories: A comparison between read-aloud and free storytelling approaches. *Educational Psychology, 38*(5), 596-616. https://doi.org/10.1080/01443410.2017.1363377

Lynch, L., Fawcett, A. J., & Nicolson, R. I. (2000). Computer-assisted reading intervention in a secondary school: an evaluation study. *British Journal of Educational Technology, 31*(4), 333-348. https://doi.org/10.1111/1467-8535.00166

MacGinitie, W. H., MacGinitie, R. K., & Company, R. P. (1989). *Gates-MacGinitie Reading Tests: Manual for scoring and interpretation*. Forms K and L, Levels 7/9 and 10/12. Riverside.

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the multimodal presence scale for virtual reality environments: a confirmatory factor analysis and item response theory approach. *Computers in Human Behavior, 72*, 276-285. https://doi.org/10.1016/j.chb.2017.02.066

Manyak, P. C., Von Gunten, H., Autenrieth, D., Gillis, C., Mastre-O'Farrell, J., Irvine-McDermott, E., Blachowicz, C. L. (2014). Four practical principles for enhancing vocabulary instruction. *The Reading Teacher, 68*(1), 13-23. https://doi.org/10.1002/trtr.1299

Massonnié, J., Bianco, M., Lima, L., & Bressoux, P. (2019). Longitudinal predictors of reading comprehension in French at first grade: Unpacking the oral comprehension component of the simple view. *Learning and Instruction, 60*, 166-179. https://doi.org/10.1016/j.learninstruc.2018.01.005

Martin, N. A., & Brownell, R. (2010). *Expressive One-Word Picture Vocabulary Test Fourth Edition (EOWPVT-4)*. Novato: Academic Therapy Publications.

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(3), 223-262. https://doi.org/10.1080/19345747.2012.755591

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly, 29*(4), 520-530. https://doi.org/10.1016/j.ecresq.2014.06.002

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior, 15*(1), 3-18. https://doi.org/10.1080/10862968309547474

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1-30. https://doi.org/10.1207/s15326950dp3801_1

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, 36*(2), 222-233. https://doi.org/10.3758/BF03195567

Meara, P., & Wolter, B. (2004). V-Links: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup & B. Henriksen (Eds.), *Angles on the English Speaking World* (pp. 85-96). Museum Tusculanum Press.

Mize, M. K., Park, Y., & Moore, T. (2018). Computer-assisted vocabulary instruction for students with disabilities: Evidence from an effect size analysis of single-subject experimental design studies. *Journal of Computer Assisted Learning, 34*(6), 641-651. https://doi.org/10.1111/jcal.12272

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*(3), 315-340. https://doi.org/10.1177/109821400302400303

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of Reading Research* (pp. 269-284). Muhwah, N. J.: Lawrence Erlbaum Associates.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237. https://doi.org/10.2307/1162893

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 63*(1), 59-81. https://doi.org/10.1353/cml.2006.0049

National Reading Panel (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.

Neumann, M. M., & Neumann, D. L. (2014). Touch screen tablets and emergent literacy. *Early Childhood Education Journal, 42*(4), 231-239. https://doi.org/10.1007/s10643-013-0608-3

Oakhill, J., Cain, K., & Elbro, C. (2019). Reading comprehension and reading comprehension difficulties. In D. A. Kilpatrick, R. M. Joshi & R. K. Wagner (Eds.), *Reading Development and Difficulties: Bridging the Gap Between Research and Practice* (pp. 83-115). Cham, Switzerland: Springer International Publishing.

Oakhill, J., Cain, K., McCarthy, D., & Field, Z. (2012). Making the link between vocabulary knowledge and comprehension skill. In A. Britt, S. Goldman & J.-F. Rouet (Eds.), *From Words to Reading for Understanding* (pp. 101-114). Routledge.

Ordóñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94*(4), 719. https://doi.org/10.1037/0022-0663.94.4.719

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing, 23*(2), 189-208. https://doi.org/10.1007/s11145-008-9159-1

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554-566. https://doi.org/10.1037/0022-0663.98.3.554

Ouellette, G., & Shaw, E. (2014). Oral vocabulary and reading comprehension: An intricate affair. *L'Année Psychologique, 114*(4), 623-645. https://doi.org/10.4074/S0003503314004023

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9-29. https://doi.org/10.18806/tesl.v11i1.623

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282-296. https://doi.org/10.1598/RRQ.42.2.4

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring Yes-No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing, 29*(4), 489-509. https://doi.org/10.1177/0265532212438053

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23-33. https://doi.org/10.1037/0022-0663.94.1.23

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L.T. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 67-86). John Benjamins Publishing.

Perfetti, C., McKeown, M., & Kucan, L. (2010). Decoding, vocabulary, and comprehension. In M. G. McKeown & L. Kucan (Eds.), *Bringing Reading Research to Life* (pp. 291-303). Guilford Press.

Potocki, A., Ecalle, J., & Magnan, A. (2016). Early cognitive and linguistic profiles of different types of 7-to 8-year-old readers. *Journal of Research in Reading, 40*, 125-140. https://doi.org/10.1111/1467-9817.12076

Protopapas, A., & Skaloumbakas, C. (2007). Traditional and computer-based screening and diagnosis of reading disabilities in Greek. *Journal of Learning Disabilities, 40*(1), 15-36. https://doi.org/10.1177/00222194070400010201

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308. https://doi.org/10.3138/cmlr.56.2.282

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52. https://doi.org/10.1191/0265532204lt273oa

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86*(1), 159-175. https://doi.org/10.1111/cdev.12292

Read, J. (2000). *Assessing Vocabulary*. Cambridge university press Cambridge.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined. In P. Bogaards & B. Laufer-Dvorkin (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 209-227). John Benjamins Publishing.

Richard, G. J., & Hanner, M. A. (2005). *Language Processing Test-3.* East Moline, IL: LinguiSystem.

Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading, 17*(6), 415-434. https://doi.org/10.1080/10888438.2013.764879

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86*(1), 54-64. https://doi.org/10.1037/0022-0663.86.1.54

Rossbach, H. G., Tietze, W., & Weinert, S. (2005). *Peabody Picture Vocabulary Test* (German research version of the test by LM Dunn & LM Dunn from 1997). Bamberg, Germany: Freie Universität Bamberg.

Rowsell, J., & Wohlwend, K. (2016). Free play or tight spaces? Mapping participatory literacies in Apps. *The Reading*

Teacher, 70(2), 197-205. https://doi.org/10.1002/trtr.1490
Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education, 58*(1), 136-153. https://doi.org/10.1016/j.compedu.2011.07.017
Scalise, K., & Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing ''intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning and Assessment, 4*(6). https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1653/1495
Schaefer, B., Bowyer-Crane, C., Herrmann, F., & Fricke, S. (2015). Development of a tablet application for the screening of receptive vocabulary skills in multilingual children: A pilot study. *Child Language Teaching and Therapy*, 1-13. https://doi.org/10.1177/0265659015591634
Schatz, P., & Browndyke, J. (2002). Applications of computer-based neuropsychological assessment. *The Journal of Head Trauma Rehabilitation, 17*(5), 395-410. https://journals.lww.com/headtraumarehab/Abstract/2002/10000/Applications_of_Computer_based_Neuropsychological.3.aspx
Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913-951. https://doi.org/10.1111/lang.12077
Schmitt, N., & Schmitt, D. R. (1993). Identifying and assessing vocabulary learning strategies. *Thai TESOL Bulletin, 5*(4), 27-33. https://eric.ed.gov/?id=ED380994
Schrank, F., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson Tests of Achievement-IV*. Riverside Publishing.
Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children: The impact of schooling. *Reading and Writing, 25*(8), 1947-1971. https://doi.org/10.1007/s11145-011-9308-9
Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics, 19*(1), 53-67. https://doi.org/10.1017/S0142716400010572
Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language, 24*(1), 123-138. https://doi.org/10.1017/S0305000996003005
Sénéchal, M., Thomas, E., & Monker, J.-A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87*(2), 218-229. https://doi.org/10.1037/0022-0663.87.2.218
Sénéchal, M., Whissell, J., Bildfell, A., Cain, K., Compton, D., & Parrila, R. (2017). Starting from home: Home literacy practices that make a difference. In K. Cain, D. L. Compton & R. K. Parrila (Eds.), *Theories of Reading Development* (pp. 383-408). John Benjamins Publishing.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795
Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *The Elementary School Journal, 108*(2), 97-113. https://doi.org/10.1086/525549
Singleton, C. (2001). Computer-based assessment in education. *Educational and Child Psychology, 18*(3), 58-74. http://lucid-research.net/documents/research/paper_jrnl_EducationalAndChildPsychology01_ComputerBasedAssessmentInEducation.pdf
Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment, 17*(4), 396-408. https://doi.org/10.1037/1040-3590.17.4.396
Snow, C. E. (1990). The development of definitional skill. *Journal of Child Language, 17*(3), 697-710. https://doi.org/10.1017/S0305000900010953
Spencer, E. J., Goldstein, H., & Kaminski, R. (2012). Teaching vocabulary in storybooks: Embedding explicit vocabulary instruction for young children. *Young Exceptional Children, 15*(1), 18-32. https://doi.org/10.1177/1096250611435367
Stetter, M. E., & Hughes, M. T. (2011). Computer assisted instruction to promote comprehension in students with learning disabilities. *International Journal of Special Education, 26*(1), 88-100. https://eric.ed.gov/?id=EJ921192
Stobart, G. (2008). *Testing Times: The Uses and Abuses of Assessment*. Routledge.
Stubbe, R. (2012). Do pseudoword false alarm rates and overestimation rates in Yes/No vocabulary tests change with Japanese university students' English ability levels? *Language Testing, 29*(4), 471-488. https://doi.org/10.1177/0265532211433033
Sung, Y.-T., Chang, K.-E., & Huang, J.-S. (2008). Improving children's reading comprehension and use of strategies through computer-based strategy training. *Computers in Human Behavior, 24*(4), 1552-1571. https://doi.org/10.1016/j.chb.2007.05.009
Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 381-398. https://doi.org/10.1207/

s1532799xssr1004_3
Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computers & Education, 56*(4), 1032-1044. https://doi.org/10.1016/j.compedu.2010.11.017
Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading, 32*(4), 383-401. https://doi.org/10.1111/j.1467-9817.2009.01401.x
Timmers, C., & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education, 56*(3), 923-930. https://doi.org/10.1016/j.compedu.2010.11.007
Tseng, W.-T. (2016). Measuring English vocabulary size via computerized adaptive testing. *Computers & Education, 97*, 69-85. https://doi.org/10.1016/j.compedu.2016.02.018
Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics, 27*(1), 78-102. https://doi.org/10.1093/applin/ami046
Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*(5), 453-466. https://doi.org/10.1177/0022219411432685
Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of Feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475-511. https://doi.org/10.3102/0034654314564881
Van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education, 58*(1), 263-272. https://doi.org/10.1016/j.compedu.2011.07.020
Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*(1), 8-25. https://doi.org/10.1080/10888438.2011.536125
Vidal-Abarca, E., Martinez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior Research Methods, 43*(1), 179-192. https://doi.org/10.3758/s13428-010-0032-1
Vloedgraven, J., & Verhoeven, L. (2009). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences, 19*(2), 161-169. https://doi.org/10.1016/j.lindif.2008.09.005
Walker, M. (2015). The cure for early grades assessment difficulties? Take a tablet. *International Developments, 5*(5), 3. https://research.acer.edu.au/intdev/vol5/iss5/3/
Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments A Meta-analysis of testing mode effects. *Educational and Psychological Measurement, 68*(1), 5-24. https://doi.org/10.1177/0013164407305592
Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93*(2), 243. https://doi.org/10.1037//0022-O663.93.2.243
Wasik, B. A., Hindman, A. H., & Snell, E. K. (2016). Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly, 37*, 39-57. https://doi.org/10.1016/j.ecresq.2016.04.003
Watkins, R. V., & DeThorne, L. S. (2000). Assessing children's vocabulary skills: From word knowledge to word-learning potential. *Seminars in Speech and Language, 21*, 235-246. https://doi.org/10.1055/s-2000-13197
Webb, S., & Chang, A. C.-S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*(4), 651-675. https://doi.org/10.1017/S0272263114000606
Wechsler, D. (1991). *Manual for the Wechlser Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.
Wechsler, D. (2003). *Manual for the Wechlser Intelligence Scale for Children* (4th ed.). San Antonio, TX: The Psychological Corporation.
Wechsler, D. (2005a). *Manuel de l'échelle d'intelligence de Wechsler pour enfants 4th édition* [*Manual for the Wechlser Intelligence Scale for Children IV*]. Paris: ECPA.
Wechsler, D. (2005b). *Wechsler Intelligence Scale for Children Fourth Edition-Spanish*. San Antonio, TX: Harcourt Assessment.
Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). Pearson. https://doi.org/10.1037/t15171-000
Wiig, E. H., & Secord, W. (1992). *TOWK: Test of Word Knowledge*. Toronto, Ontario, Canada: Psychological Corporation.
Wiig, E., Secord, W., Semel, E., Boulianne, L., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales: Version pour francophones du Canada* [*Clinical evaluation of language fundamentals: French*

Canadian version]. Toronto, Ontario, Canada: Pearson Canada Assessment.
Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical Evaluation of Language Fundamentals: CELF-5*. Pearson.
Williams, K. T., & Pearson Education, I. (2018). *EVT-3: Expressive Vocabulary Test-III*. Pearson Assessments.
Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183. https://doi.org/10.1207/s15324818ame1802_2
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement-III*. Riverside Publishing.

Appendix A

Table 1. Tools of Vocabulary Assessment According to Vocabulary's Dimension Assessed, Tasks, Populations, and Language

Tests or subtests	Authors (years)	Dimensions assessed	Tasks	Age range	Language
BILO* Vocabulary subtest	Khomsi et al. (2007)	Breadth, Receptive	Matching word with multiple choice pictures	3.0 to 15.11 years	French
BPVS-3	Dunn et al. (2009)	Breadth, Receptive	Matching word with multiple choice pictures	3.0 to 16.11 years	English
CELF-4 Word Classes subtest	Kort et al. (2008)	Depth, Expressive	Defining target words and indicating connection between words	5.0 to 16+ years	Dutch
CEFL4 Word Classes subtest	Wiig et al. (2009)	Depth, Expressive	Defining target words and indicating connection between words	4.0 to 16.11 years	French-Canadian
CELF-5* Word Classes subtest	Wiig et al. (2013)	Depth, Expressive	Indicating word related by semantic features	5.0 to 21.11 years	English
DIBELS-6* Word Use Fluency subtest	Good and Kaminski (2002)	Depth, Expressive	Correctly use words in sentences	3.0 to 11.11 years	English
DIBELS* Word Use Fluency subtest	Dufour-Martel and Good (2009)	Depth, Expressive	Correctly use words in sentences	3.0 to 11.11 years	French
EOWPVT	Groupe coopératif en orthophonie-Région Laval, Laurentides, Lanaudière (1995)	Breadth, Expressive	Naming of pictures	2.0 to 18.11 years	French
EOWPVT	Brownell (2001)	Breadth, Expressive	Naming of pictures	4.0 to 12.11 years	Spanish bilingual/ English
EOWPVT-4	Martin and Brownell (2010)	Breadth, Expressive	Naming of pictures	2.0 to 70+ years	English
EVT-3*	Williams and Pearson Education (2018)	Breadth, Expressive	Naming of pictures	2.6 to 90+ years	English
Gates McGinitie. Vocabulary subtest	MacGinitie et al. (1989)	Depth, Receptive	Choose the word of the sentence the closest to the meaning of a target word	10.0 to 18.11 years	English
KABC Vocabulary subtest	Kaufman and Kaufman (2008)	Breadth, Receptive	Matching word with multiple choice pictures	3.0 to 12.11 years	French
KABC-2 Vocabulary subtest	Kaufman et al. (2014)	Breadth, Receptive	Matching word with multiple choice pictures	3.00 to 18.11 years	German
KABC-2 Vocabulary subtest	Kaufman et al. (2018)	Breadth, Receptive	Matching word with multiple choice pictures	3.00 to 18.11 years	English
L2MA-2* Vocabulary subtest	Chevrie-Muller et al. (2011)	Breadth, Expressive	Naming of pictures and categorization of words	7.0 to 11.11 years	French

Tests or subtests	Authors (years)	Dimensions assessed	Tasks	Age range	Language
LPT-R Multiple Meanings and Attributes subtests	Richard and Hanner (2005)	Depth, Expressive and Receptive	Defining words in different contexts Describing attributes of names	5.0 to 11.11 years	English
PPVT	Dunn et al. (1986)	Breadth, Receptive	Matching word with multiple choice pictures	2.6 to 17.11 years	Spanish, bilingual English
PPVT	Dunn et al. (1993)	Breadth, Receptive	Matching word with multiple choice pictures	2.6 to 18.00 years	French
PPVT	Rossbach et al. (2005)	Breadth, Receptive	Matching word with multiple choice pictures	4.0 to 8.11 years	German
PPVT-IV*	Dunn and Dunn (2007)	Breadth, Receptive	Matching word with multiple choice pictures	2.6 to 90+ years	English
TACL-Revised Words classes and relations subtest	Carrow-Woolfolk (1985)	Breadth, Receptive	Selecting the picture that represent the meaning of words or sentence	3.0 to 9.11 years	English
TACL Words classes and relations subtest	Carrow-Woolfolk (1985)	Breadth, Receptive	Selecting the picture that represent the meaning of words or sentence	3.0 to 21.11 years	French, Canadian
TOWK Expressive and Receptive vocabulary	Wiig and Secord (1992)	Depth and Breadth, Expressive	Defining a target word; Matching expressions with multiple choice pictures; Matching word with multiple choice synonyms	5.0 to 17.11 years	English
WASI Vocabulary subtest	Wechsler (2011)	Depth, Expressive	Defining a target word	6.0 to 90.11years	English
WISC-IV Expressive Vocabulary subtest	Wechsler (2003)	Depth, Expressive	Naming of pictures Defining a target word	6.0 to 16.11 years	English
WISC-IV Expressive Vocabulary subtest	Wechsler (2005a)	Depth, Expressive	Naming of pictures Defining a target word	6.0 to 16.11 years	French
WJ-III Picture Vocabulary subtest	Woodcock et al. (2001)	Breadth, Expressive and Receptive	Naming of pictures, Identifying pictures objects	2.0 to 90+ years	English
WJ-IV Picture Vocabulary subtest	Schrank et al. (2014)	Breadth, Expressive and Receptive	Naming of pictures, Identifying pictures objects	2.0 to 90 + years	English

Notes: BILO = Bilan Informatisé de Language Oral [Computerized Checkup for oral language]; BPVS = British Picture Vocabulary Scale; CELF = Clinical Evaluation of Language Fundamentals; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; EOWPVT = Expressive One-Word Picture Vocabulary test; EVT = Expressive Vocabulary Test; KABC = Kaufman Assessment Battery for Children; L2MA-2 = Batterie Language Oral, Language Ecrit, Mémoire, attention [Oral language, written, language, memory, attention]; LPT = Language Processing Test; PPVT = Peabody Picture Vocabulary Test; TACL = Test for Auditory Comprehension of language; TOWK = Test of Word knowledge; WASI = Wechsler Abbreviated Scales of Intelligence; WISC = Wechsler Intelligence Scale for Children; WJ = Woodcock Jonhson. * indicates that a digital version can be available

Appendix B

Table 1. Overview of Studies Characteristics Included in the Current Review for Instruction of Vocabulary

Authors	Participants characteristics	Objectives	Variables, language skills	Language	Research design	Main Findings and Effect size
Beck and Mckeown (2007) (Study 1)	98 kindergarteners and first-graders	Examining the effect of vocabulary instruction (sophisticated words)	Text Talk, receptive Vocabulary skills, prior knowledge	English	Between-subjects, quasi-experimental, pretest and posttest control group design	Experimental group learned more words for kindergartners (d = 1.17), for first-graders (d = 0.744)
Beck and Mckeown (2007) (Study 2)	76 kindergarteners and first-graders	Examining the effect of two amounts of instruction (More Rich vs. Rich instruction)	Text Talk, receptive Vocabulary skills, prior knowledge, Verbal skills	English	Within-subject, quasi-experimental, pretest and posttest design	Kindergartners: Benefit from more rich instruction for verbal test (d = 2.09) and for pictures test (d = 2.71) First grade: Benefit from more rich instruction for verbal (d = 2.09) and for pictures (d = 2.71) tests
Beck et al. (1982)	66 fourth-graders	Examining the effect of vocabulary instruction (vocabulary vs. language arts)	Reading and Vocabulary Skills; Semantic decision latency, sentence verification latency, story recall	English	Within-subject and between-subjects comparison of experimental and control group	Experimental group better performances for vocabulary tests, but no for semantic decision or sentence verification (marginal effect), or story recall
Beschorner and Hutchison (2013)*	35 Four- and five-year-old children	Examining the role of iPads as instructional tools in two preschool classrooms	Writing, speaking, listening and print awareness apps	English	Within-subject and observational design	Develop digital print awareness Emergent writing skills with digital technology. Benefit from connecting, reading, writing, listening and speaking. Social learning with more interaction between children
Biemiller and Boote (2006) (Study 1)	43 kindergarteners, 37 first graders, 32 second graders	Examining the effect of pretesting, reading books, and word explanations on word meaning acquisition during instruction	Story books, expressive word knowledge	English	Within-subject, between-subject, quasi-experimental, pretest and posttest control group design	Higher scores on posttest (d = 1.21). Effect of grade, and gender with higher performances for girls on gains. Higher scores for instructed words (d = 0.53), with additional gain to repeated reading. Effect of reading book times according to grades, with benefit for kindergarten, and Grade 1
Biemiller and Boote (2006) (Study 2)	28 kindergarteners, 37 first graders, 42 second graders	Examining the effect of instruction procedures (intensive, word and transfer) on the percentage and number of word meanings acquired	Story books, expressive word knowledge	English	Within-subject, between-subject, quasi-experimental, pretest and posttest control group design	Effect of pretests and posttest-delayed on gains (d = 2.97), larger for Grade 1. Gains between pretest and posttest (d = 2.30), larger for Grade 1. Gains between posttest and posttest-delayed (d = 0.26). Benefit from repeated oral reading of stories combined with explanations of words and reviews of words explained

Authors	Participants characteristics	Objectives	Variables, language skills	Language	Research design	Main Findings and Effect size
Chen and Sun (2012)*	160 fifth-graders	Examining the effects of multimedia material (static text, video, animated and interactive) on learning performance and emotions according to individual cognitive styles (verbal, visual)	Learners' emotions, learners' performances multiple-choice tests	Mandarin	Between-subject, quasi-experimental, pretest and posttest design	Video-based multimedia: Best learning performance and most positive emotion for verbalizers. Video-based and animated/ interactive multimedia: More appropriate for visualizers. Partial correlation of negative emotion on learning performances for visualizers
Christ and Chiu (2018)	56 kindergarteners	Exploring the effect of words presentation without direct instructions (control, read-aloud, teacher conversations, or both) on depth word knowledge	Receptive vocabulary knowledge, books story	English	Between-subject, quasi-experimental, pretest and posttest design	Benefit from both read-aloud and teacher conversations for depth word knowledge
Couse and Chen (2010)*	41 preschoolers	Exploring the viability of tablet computer by drawing in early education	Drawing apps on tablet with stylus	English	Explanatory research with both a multiple single-subject and qualitative research design	Increased engagement, interest and ability to use tablets with age and through sessions
Coyne et al. (2009)	42 kindergarteners	Comparing the effect of embedded (breadth), extended (depth) and incidental instructions of vocabulary knowledge	Storybook read-aloud, expressive definition, context/full knowledge, context/partial knowledge, receptive definition, receptive vocabulary skills	English	Within-subject, experimental, posttest and posttest-delayed design	Expressive definition: Benefit from extended instruction compared to embedded (d = 1.34), which in turn higher than for incidental (d = 0.87). Receptive definition: benefit from extended instruction (d = 0.70) which in turn higher than for incidental (d = 0.24)
Coyne et al. (2010)	80 kindergarteners	Investigating the direct and extended vocabulary instruction on target words and transfer of generalized language and literacy	Interactive story read aloud, receptive vocabulary skills, listening comprehension, expressive and depth vocabulary knowledge, metalinguistic awareness	English	Between-subject, quasi-experimental control group design	Initial receptive vocabulary knowledge predicts target word definition (d between 1.06 to 2.44) Benefit from direct and extended vocabulary instruction (d = 1.71) Moderate effect of transfer for listening comprehension (d = 0.42)
Coyne et al. (2004)	64 kindergarteners	Investigating a storybook intervention with explicit vocabulary instruction on vocabulary development in children at risk of reading difficulty	Receptive and expressive vocabulary skills, prior knowledge	English	Between-subject, quasi-experimental, pretest and posttest, control group design	Benefit from explicitly taught vocabulary for children at risk with low receptive vocabulary
Cunningham and Stanovitch (1991)	34 fourth-, 33 fifth- and 67 sixth-graders	Examining the contribution of print exposure to children's verbal abilities	Exposure to print, general cognitive abilities, phonological coding, spelling, vocabulary knowledge, verbal fluency, receptive vocabulary skills, general information	English	Between-subject design	Print exposure contributes to the development of verbal abilities
Elley (1989) (Study 1)	157 7-years-old children	Examining the effect of stories read aloud by teacher (without explanation) on vocabulary growth	Story Book, multiple choice vocabulary test (matching picture, synonym)	English	Between-subject, pretest and posttest design	Increased vocabulary after listening to story book (3 times), especially for low vocabulary groups

Authors	Participants characteristics	Objectives	Variables, language skills	Language	Research design	Main Findings and Effect size
Fernández-López et al. (2013)*	34 students from 4 to 20 years old	Examining the effect of mobile learning platform on learning skills in student with special education needs	Mobile apps: Language, math, environmental, awareness, autonomy and social	Spanish	Between-subject, pretest and posttest design	Benefit from mobile learning platform for basic learning skills (association, exploration, puzzle, sorting)
Herman et al. (1987)	309 eighth-graders	Investigating the effect of different types of text features (i.e., macrostructure, microstructure, concept elaborations) on reader's incidental acquisition of vocabulary knowledge	Texts, comprehensive skills, expressive and receptive, vocabulary skills	English	Within-subject, between subject design	Benefit from conceptually elaborated version text for word knowledge
Hutchison et al. (2012)*	23 fourth-graders	Exploring and understand the viability of using iPads to support and enhance literacy instruction	Literacy activities on apps (e.g., reading skills, drawing)	English	Within-subject, exploratory design	Advantages of iPads: Creativity in idea presentation, engagement in activities, individualization of book selection. Development of digital literacy skills
Jenkins et al. (1984)	112 fifth-graders	Examining the acquisition of vocabulary knowledge through incidental learning of word meanings from context	Prior knowledge, reading ability, vocabulary knowledge	English	Within-subject and between-subject, quasi-experimental design	Benefit from higher reading ability and prior knowledge for vocabulary knowledge. Effect of the number of exposure context on vocabulary knowledge
Johnson et al. (1982)	36 fourth to six grade classes	Comparison of the instructional strategies of semantic mapping and semantic feature analysis with a contextual approach of vocabulary acquisition	Word knowledge, comprehensive test	English	Within and between experimental design	Benefit from both semantic features analysis and semantic mapping compared to context for general vocabulary acquisition
Johnson-Glenberg (2007) (Study 1)*	6 seventh-, 14 sixth-graders	Investigating a Web-based application for comprehension instruction (3D-Readers) on poor comprehenders	Science text, vocabulary knowledge, reading comprehension, metacognitive strategy, prior knowledge	English	Within-subject design	Benefit from 3D-Readers for constructing answers to open-ended questions over eight sessions
Johnson-Glenberg (2007) (Study 2)*	11 third- to eighth-graders	Investigating a Web-based application for comprehension instruction (3D-Readers) on students with attention deficit disorder/attention deficit hyperactivity disorder	Science text, vocabulary knowledge, reading comprehension, metacognitive strategy, prior knowledge	English	Within-subject design	Benefit from vocabulary, and in self-reported metacognitive strategy use after six sessions
Johnson-Glenberg (2007) (Study 3)*	37 fourth- to seventh-graders	Investigating a Web-based application for comprehension instruction (3D-Readers) on summer school students	Science text, vocabulary knowledge, reading comprehension, metacognitive strategy, prior knowledge	English	Within-subject design	Benefit from vocabulary skill, and in the quality of questions generated over four sessions

Authors	Participants characteristics	Objectives	Variables, language skills	Language	Research design	Main Findings and Effect size
Justice et al. (2005)	57 kindergartners	Examining the influence of storybook reading exposure on the acquisition of vocabulary for at-risk children according to individual differences and elaboration of word in context	Storybook, priori vocabulary knowledge	English	Within-subject, between-subjects, pretest and posttest, quasi-experimental comparison group design	Elaborated words: Benefit from repeated storybook reading for the acquisition of vocabulary (d = 1.22); gain more important for low vocabulary skills (d = 1.34). Non-elaborated words: no effect of storybook reading (d = 0.53)
Lenhart et al. (2018) (study 1)	83 preschoolers	Examining the effect of story delivery (read aloud vs. told freely) on the acquisition of vocabulary	Receptive vocabulary, phonological working memory, speech comprehension	German	Within-subjects, pretest and posttest, experimental design	Acquisition of vocabulary between pretest and posttest (d = 0.37). No effect of story delivery
Lenhart et al. (2018) (study 2)	48 preschoolers	Examining the effect of story reading delivery (read aloud vs. told freely) on the acquisition of vocabulary according to age	Receptive vocabulary, phonological working memory, speech comprehension	German	Within-subjects, pretest and posttest, quasi-experimental design	Younger children: Benefit from reading aloud story for the acquisition of vocabulary (d = 0.13). Oldest children: Benefit from told freely story for the acquisition of vocabulary (d = 0.20)
Loftus and Coyne (2013) (Study 1)	124 kindergarteners	Developing the strong intervention according to the number of storybook reading, types of intervention activities.	Storybooks, word knowledge, listening comprehension, metalinguistic awareness, priori vocabulary knowledge	English	Within-subject, between-subject, Quasi-experimental design	Benefit from intervention (d = 1.71), more important for initial high vocabulary skills Strong effect of intervention on transfer measures (ds = 0.73-0.81)
Loftus and Coyne (2013) (Study 2)	43 kindergarteners	Examining the effectiveness of whole-class vocabulary instruction according to a multi-tier approach in at-risk and not-at risk students	Storybooks, word recognition, context questions, expressive definition	English	Within-subject, between-subject, Quasi-experimental design	Benefit from Tier 2 vocabulary intervention for measures of literacy skills (ds = 0.40-0.69) Both Tier 1 ad Tier 2 intervention in at-risk student decrease differences with not-at risk student (ds = 0.70-0.72)
Lynch et al. (2000)*	8 children in secondary school	Evaluating the effectiveness of a computer-based literacy support system (RITA) in children with disadvantaged literacy skills	Reading speed and accuracy, reading comprehension, Spelling, vocabulary	British	Within-subject, pretest and posttest, quasi-experimental, group control design	Benefit from training for reading (d = 0.49), spelling (d = 0.18), and literacy standard scores (d = 0.39) Increase of enthusiasm and engagement in children
McKeown and Beck (2014)	131 kindergartners	Examining the effects of two approaches to vocabulary instruction, repetition and interactive, and a control group in children	Storybook, meaning recognition, comprehension, production, context integration	English	Within-subject, experimental, pretest and posttest design	Benefit from repetition and interactive conditions for recognition of word meanings (ds = 0.35, 0.44), context integration (ds = 0.27, 0.38), and production (ds= 0.44, 0.70). Benefit from interactive instruction for context integration (d = 0.21), and production (d = 0.26)
McKeown et al. (1983)	Fourth-graders in two schools	Investigating the relationship between vocabulary instruction and reading comprehension	Reading, vocabulary, accuracy, fluency, text comprehension	English	Within-subject, between-subject, quasi-experimental, pretest and posttest design	Benefit from instruction for vocabulary knowledge, accuracy, fluency, and comprehension

Authors	Participants characteristics	Objectives	Variables, language skills	Language	Research design	Main Findings and Effect size
Nagy et al. (1987)	157 third-, 100 fifth-, and 160 seventh-graders	Investigating incidental learning of word meanings from context during normal reading	Expository or narrative texts, vocabulary knowledge, reading skills	English	Within-subjects, between-subjects, quasi-experimental design	Increase of word knowledge for the read text. Benefit from easy text on word knowledge
Ordonez et al. (2002)	38 fourth- and 50 fifth-graders	Examining the relationship between paradigmatic and syntagmatic word knowledge in bilingual children	Vocabulary knowledge (depth and breadth), communicative skills, paradigmatic or superordinate knowledge, syntagmatic or nonhierarchical knowledge	Spanish, English	Within-subject design	Paradigmatic knowledge in Spanish predicts paradigmatic knowledge in English, when breadth in both languages is controlled. Communicative skills both in English and Spanish only when breadth in both languages is controlled. Deep word knowledge depends on English vocabulary knowledge
Penno et al. (2002)	47 first-graders	Evaluating the effect of repeated listening to stories and teacher's explanation on children's vocabulary growth	Stories, vocabulary knowledge, generalization words	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from listening stories for vocabulary growth, with both frequency and explanation from teachers, especially in children with low vocabulary abilities
Robbins and Ehri (1994)	51 kindergarteners	Examining the effect of listening to stories on kindergarteners' vocabulary growth	Stories, vocabulary knowledge, word heard or not heard in story	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Better knowledge from words heard in a story than words unheard. Vocabulary gains more important for children with high-vocabulary knowledge. Gain with at least 4 exposures to word for vocabulary growth
Sénéchal (1997)	30 three- and 30 four-year-old children	Examining the effect of didactic techniques used during storybook reading (single reading, repeated-reading, questioning) on young children's vocabulary growth	Vocabulary skills, prior knowledge	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from repeated listening to storybook for both receptive and expressive vocabulary growth. Answering questions better for expressive vocabulary than for receptive vocabulary
Sénéchal et al. (1995) (Study 1)	32 4-year-old children	Assessing how children's vocabulary skills influence vocabulary growth according to listening to stories (listening, labeling)	Prior knowledge, storybook, comprehensive vocabulary, production vocabulary, home literacy, socioeconomic status	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from answering to questions during reading for comprehensive vocabulary and production vocabulary. Benefit from prior knowledge for production vocabulary
Sénéchal et al. (1995) (Study 2)	48 4-year-old children	Assessing how children's vocabulary skills influence vocabulary growth according to conditions of listening to stories (listening, pointing, labeling	Prior knowledge, storybook, comprehensive vocabulary, production vocabulary, home literacy, socioeconomic status	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from prior knowledge for understanding of new words Benefit from active participation during listening to story for vocabulary growth
Silverman (2007) (Study 1)	94 kindergarteners	Comparing 3 approaches to teaching vocabulary during storybook reading at the beginning of instruction: contextual vs. analytic vs. anchored	Storybook, socioeconomic status, prior vocabulary knowledge (expressive and receptive), general vocabulary, home language	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from both analytic and anchored for receptive vocabulary (ds = 0.67, 1.02); for oral vocabulary (ds = 0.85, 1.19)

Authors	Participants characteristics	Objectives	Variables, language skills	Language	Research design	Main Findings and Effect size
Silverman (2007) (Study 2)	50 first-graders	Investigating the effectiveness of 3 approaches to teaching vocabulary during storybook reading after 6 months: contextual vs. analytic vs. anchored	Storybook, socioeconomic status, prior vocabulary knowledge (expressive and receptive), general vocabulary, home language	English	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from both analytic and anchored for oral vocabulary (*ds* = 0.58, 0.94), but no significant effect on receptive vocabulary
Sung et al. (2008)*	130 sixth-graders	Examining the effects of Computer Assisted Strategy Teaching and Learning Environment (CASTLE) on students with different reading disabilities	Reading comprehension, reading ability, narrative and expository texts, reading strategies	Mandarin	Within-subjects, between-subjects, quasi-experimental, pretest and posttest design	Benefit from CASTLE for using strategies and for comprehension only for expository texts
Wasik and Bond (2001)	124 4-years-olds children	Examining the effect of interactive book reading on language and literacy development of children from low socioeconomic status	Storybooks, receptive vocabulary skills, receptive and expressive language skills	English	Between-subjects, quasi-experimental, pretest and posttest design	Benefit from interactive book reading intervention on receptive and expressive language and literacy skills
Webb and Chang (2015)	60 sixteen-year-old students	Investigating in which English foreign language learners according to vocabulary sizes learned through reading 20 graded readers and whether vocabulary learning gains according to reading text difficulty levels	Vocabulary knowledge, prior knowledge, graded readers (Term 1 or Term 2)	English, Mandarin	Between-subjects, quasi-experimental, pretest and posttest design	Learning gain more important for high level group (*d* = 0.98) Vocabulary gain more important for high level group (*d* = 1.12) for Term 1 Vocabulary gain more important for high- and intermediate level group for Term 2