



# Language Teaching Research Quarterly

2024, Vol. 46, 297–324



## Dynamic Assessment of Vocabulary: A Systematic Literature Review

Yuwei Xia\*, Ying Luo, Xiaofei Lu

The Pennsylvania State University, United States

Received 14 April 2024

Accepted 16 November 2024

### Abstract

Grounded in Vygotsky's Sociocultural Theory, Dynamic Assessment (DA) can provide fine-tuned information regarding learners' development and opportunities to promote further development. Vocabulary knowledge, because of its complex nature, requires a more sensitive assessment, especially for second language (L2) learners. This systematic literature review investigated 14 empirical studies that used DA to address L2 vocabulary knowledge to provide a comprehensive picture of what has been done, including the target aspects, instruments, and mediational strategies to answer the questions of what kind of information DA could provide and how well it could promote vocabulary development. The findings indicate that the DA of vocabulary is still premature, and more attention is needed in developing a more encompassing instrument and mediational strategies.

**Keywords:** *Systematic Literature Review, Dynamic Assessment, Second Language Vocabulary, Mediation*

### How to cite this article (APA 7<sup>th</sup> Edition):

Xia, Y., Luo, Y., & Lu, X. (2024). Dynamic assessment of vocabulary: A systematic literature review. *Language Teaching Research Quarterly*, 46, 297-324. <https://doi.org/10.32038/ltrq.2024.46.19>

### <sup>1</sup>Introduction

The field of language education has increasingly realized the importance of assessment for its ability to provide information about learners' knowledge and ability (Lynch, 2001). Research on traditional assessments has corroborated their importance but at the same time also revealed several critical limitations. First of all, despite their psychometric soundness, traditional assessments such as standardized tests require learners to complete the tests alone with no access to other resources. Consequently, the results obtained from such assessments can only

<sup>1</sup> This paper is part of a special issue (2024, 46) entitled: In Honour of James P. Lantolf's Contributions to Sociocultural Theory, Second Language Development and Language Pedagogy (edited by Mirosław Pawlak, Zhisheng (Edward) Wen, and Hassan Mohebbi).

\* Corresponding author.

E-mail address: [yxx5304@psu.edu](mailto:yxx5304@psu.edu)

<https://doi.org/10.32038/ltrq.2024.46.19>

shed light on learners' past social and historical activities and reflect their fully developed abilities. However, these results usually serve gatekeeping purposes, such as admission, selection, or placement (Messick, 1995), to predict whether learners will succeed in the future. Using past evidence for future decisions is based on the problematic *past-to-present* (Valsiner, 2001) assumption that the future is a smooth continuation or extension of the past. As pointed out by Vygotsky and Rieber (1998), development is not linear, nor could it be predicted by the past alone (Poehner, 2008). Educational decisions made based on evidence indicating past achievement significantly hampered the consequential aspect of the construct validity of these assessments (Messick, 1995).

Second, traditional assessments are often administered at the end of the learning cycle to show what learners can and cannot do; the timing of these assessments and their backward look nature do not give insight into learning or development. In addition, because of the ubiquitous influence of tests and assessments in our current educational system for evaluation purposes, educators are concerned with "teaching to the test," focusing heavily on materials that will be tested along with test-taking strategies (Poehner, 2008). Therefore, the relationship between assessment and instruction is separate at best and antagonistic at worst. After Messick (1995) pointed out the washback effect of tests on teaching and learning, researchers, educators, and policymakers have had intense discussions over the relationship between assessment and instruction. In addressing the limitations of traditional assessment, many alternative assessment frameworks, such as learning-oriented assessment (Carless, 2007; Purpura, 2004), have been proposed to reconcile the tension between assessment and instruction. Among these various attempts, Dynamic Assessment (DA) has attracted increasing scholarly interest in the language testing field as an alternative assessment procedure for its many unique advantages (Poehner & Wang, 2021).

Vocabulary, as an important aspect of language, has regained researchers' attention after being downplayed for several decades (Meara, 1997). An increasing number of scholars pointed out that the complexity of vocabulary knowledge calls for a more comprehensive and sensitive measurement (Read, 2000; Schmitt et al., 2020). While DA has been extensively implemented and researched for many different aspects of language, DA of vocabulary remained a scarcely researched topic with only a handful of projects focusing on adult language learners' vocabulary development. In this review, we aim to examine the current literature on DA of vocabulary, with a focus on the aspects of vocabulary knowledge targeted and the mediational moves adopted, to understand the effect of DA on assessing and promoting vocabulary knowledge development and make suggestions in terms of assessment instruments and assessment procedures.

## **Background**

### *Dynamic Assessment*

Rooted in both Vygotsky's (1987) Sociocultural Theory (SCT) and Feuerstein's Mediated Learning Experience (MLE) framework (Feuerstein et al., 1979), DA challenges traditional assessment procedures by considering mediation, dialectically integrates instruction and assessment in a way that provides insights into learners' current development as well as their emerging and not yet fully matured abilities.

According to Vygotsky (1987), the development of human higher-level functions occurs during social interactions, through the process of internalization, and is then transformed into cognitive abilities. Therefore, cognitive development is neither innate nor static, but instead is acquired and dynamic—thus, it is much more complicated than what is revealed by traditional assessments. Learners' independent problem-solving ability, as manifested in traditional assessments, reveals only the fully internalized aspects of the learners' cognitive functions. However, to understand learners' development more accurately, it is critical to take their emerging and not yet fully matured abilities into consideration. For example, in a traditional assessment setting, two learners might both answer the same vocabulary question incorrectly, suggesting a lack of sufficient knowledge. However, further interaction could reveal different reasons for their mistakes: one learner may have never encountered the target word before, while the other may have some familiarity with the target word but not enough knowledge about the specific aspect being tested.

Instead of leaving learners to complete the assessment tasks independently, various kinds of mediation are provided per learners' needs when difficulties manifest during DA procedures. Learners' reactions and responses to the support or mediation provided during the assessment procedure while dealing with difficult questions can provide more fine-tuned information regarding their not-yet-matured abilities; in addition, analyzing their responses can help teachers plan and implement future instruction to further promote learning. Because of this dual function, DA can provide both summative and formative information for diagnosis and prognosis purposes.

### *Zone of Proximal Development*

Vygotsky and Cole (1978) believed that development emerges through guided and supported activities via the use of culturally available resources. In other words, knowledge development originates within the social environment. In addition, they argued that the development of new cognitive functions appears twice, the intermental functioning as learners interact with others and, later, the intramental functioning as they internalize their intermental functioning and can finish the task without support. Therefore, social interaction and mediation are the basis for knowledge development (Sternberg & Grigorenko, 2001). Learners' independent performance, as required by traditional assessments, represents their zone of actual development (ZAD), their fully developed and already internalized abilities of which they can exercise full self-regulation. When their performance breaks down, as indicated by their failure to answer certain items or perform certain tasks, the limits of their knowledge capacity are revealed. That is all traditional assessment could tell us. However, when taking mediation and support into consideration, we can gain much more valuable information. The zone of proximal development (ZPD) (Vygotsky & Cole, 1978) represents the gap between learners' independent abilities and what they can achieve with guidance.

As the theoretical foundation of DA, ZPD has brought several important implications. First, as many empirical studies have shown, students with the same initial ability level, as indicated by their test scores, show significantly different results after instruction. Cobelieved that these differences are due to their different ZPDs. Second, assessing learners' ZPD by investigating their reactions toward different forms of mediation can reveal detailed information about the

range of their still-developing abilities. Because of these understandings, instruction can target the learner's ZPD and promote learning more effectively.

### *Mediation*

The concept that the human mind is mediated (Lantolf, 2000) is the cornerstone of SCT and is particularly central to DA. As the defining feature of human psychology (Vygotsky, 1987), mediation is conceptualized as the instrument of cognitive change in the construct of activity and generation of higher mental processes (Donato & McCormick, 1994). Embedded in sociocultural contexts, mediation can have different forms, such as objects, symbols, and persons. Through mediation, DA challenges traditional views of assessment and dialogically integrates assessment and instruction. When difficulties arise, learners' responsiveness to mediational moves, such as reminders, leading questions, hints, provision of a model, and feedback (Haywood & Lidz, 2006), can provide insight regarding their not yet fully developed abilities and in the meantime provide information for subsequent instruction (Poehner & Wang, 2021).

Feuerstein (1979) outlined 11 attributes of mediation, among which the most important three are *intentionality*, *transcendence*, and *meaning*. *Intentionality* refers to the mediator's explicit and deliberate intention to teach. According to Lids (1991), these deliberate activities include behaviors such as "initiating, maintaining, and terminating the interaction" and "regulating and refocusing learners' attention and participation" (pp. 74-75). *Transcendence* means that mediation should help learners move beyond the task at hand and address the hypothetical "what if" question (Lidz, 1991). Feuerstein et al. (1979) pointed out that the goal of MLE is to move learners' cognitive development from the initial training phase progressively to more complex modifications that require learners to adapt what has been learned. *Meaning* emphasizes that the significance of an object or action should be clearly mediated rather than intuitively understood. As the main focus of MLE, mediation of meaning holds intentionality and transcendence together and explains what needs to be mediated and what mediation looks like. In addition, with development as its ultimate goal, mediation, manifesting in forms such as reasoning, conceptual knowledge, and most importantly, language, can regulate learners' psychological activities, thus fostering new ways of understanding, thinking, and acting (Poehner & Wang, 2021).

### *Different Approaches to DA*

Poehner and Lantolf (2005) outlined two general approaches to DA. The first is the interactionist DA, which emphasizes negotiating mediation during actual interactions with students. According to Feuerstein et al. (1988), development depends on the present and, more importantly, the appropriateness of the interaction and instruction. Interactionist DA is based on the assumption that mediation needs to be relevant to the current tasks and characterized by intentionality, reciprocity, and transcendence. As such, it should emerge during interactions and target individual learners' ZPD to provide maximum support for future development.

The interactionist DA model has shown great potential in accurately measuring learners' emerging abilities and promoting development. Meanwhile, it has also met several criticisms. First of all, practicing DA with this approach is time-consuming, labor-intensive, and sometimes not feasible (Poehner & Lantolf, 2013). Teachers or mediators need to possess

extensive content knowledge and assessment literacy, especially regarding when and how to deliver appropriate mediation during interactions. Furthermore, since mediation is tailored to individual learners or groups of learners, it can only be effectively implemented with a limited number of learners. In addition, many psychometricians contend that the individualized nature of DA practice limits the generalizability or comparability of its results among learners. However, Poehner (2008) pointed out that Vygotsky approached development from a distinct ontological perspective, and therefore, the standards used to evaluate traditional assessments may not be applicable or relevant in the context of DA (Ratner, 2006).

Another approach to DA is interventionist DA, which standardizes mediation with greater emphasis on the psychometric properties of the assessment procedure (Poehner & Lantolf, 2005). With some compromises, interventionist DA can yield readily results that are more readily comparable among learners, learner groups, and different assessment formats. A significant advantage of interventionist DA is that it allows for pre-scripted mediation, making it possible to be computerized and efficiently administered to larger groups of learners with no time and place restrictions (Poehner & Lantolf, 2013; Poehner, Zhang, & Lu, 2015; Zhang & Lu, 2019).

#### *The Nature of L2 Vocabulary Knowledge (VK)*

Vocabulary is an essential aspect of language. As Wilkins (1972) pointed out, without vocabulary, meaning cannot be communicated among language users. Many empirical studies in L2 acquisition have shown that vocabulary is closely related to various language skills, including reading (Haastrup, 2008), listening, speaking, and writing (Alderson, 2005). Therefore, measuring language learners' vocabulary knowledge (VK) is necessary for understanding their overall language proficiency.

The importance of vocabulary instruction and assessment has gained renewed attention following Nation's (1990) article highlighting the critical role of vocabulary in L2 acquisition and the different nature of first language (L1) and L2 VK. At first glance, VK may appear to be a straightforward construct; however, closer investigations have revealed its complexity. Several different ways to describe VK exist (e.g., Read, 2004; Schmitt, 2000), among which Nation's (2001) framework of VK (see Table 1) is the most widely accepted and cited. In his framework, Nation conceptualizes VK as comprising nine aspects across form, meaning, and use, addressing both receptive and productive components. In addition, VK is incremental in nature, which means that the above-mentioned aspects are not mastered simultaneously or at the same rate. Certain types of word knowledge develop before others, and learners' understanding of a vocabulary item deepens with multiple exposures (Nation, 2021). As Haastrup (2008) illustrated, learners' vocabulary size improves over time for both L1 and L2, suggesting that vocabulary size is neither stable nor static. These characteristics indicate that VK development is a complicated and gradual process. Therefore Schmitt (2014) argued that learners' VK cannot be measured through a simple know/not know dichotomy. However, as Read and Chapelle (2001) claimed, current measures of vocabulary size and growth still largely rely on counting, classifying, and assessing knowledge of individual word forms, highlighting the need for more sensitive VK measurement methods (Read, 2004; Schmitt, 2014).

**Table 1***Description of Vocabulary Knowledge* (from Nation, 2001, p.27)

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Association	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does this word occur?
		P	In what patterns must we use this word?
	Collocation	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use	R	When, where, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

*Note.* R = receptive knowledge, P = productive knowledge.

Researchers generally agree that in L1 acquisition most words are acquired incidentally (Nagy et al., 1985; Nation, 2001). However, research in L2 acquisition indicates that incidental vocabulary learning is neither efficient nor effective (Cobb, 2007; Webb, 2008) and that deliberate learning is crucial. Therefore, DA, because of its ability to provide more nuanced information and integrate instruction, holds significant potential for both vocabulary assessment and instruction.

### *Previous Literature Reviews*

Recent years have seen many DA attempts in L2 acquisition. Several literature reviews have synthetically investigated the effect of DA and its mediation on promoting L2 learners' listening (e.g., Ableeva, 2018; Poehner et al., 2015), speaking (e.g., Hill & Dabet, 2009), reading (e.g., Davin et al., 2014; Pishghadam et al., 2011), writing (e.g., Poehner & Yu, 2022; Shrestha & Coffin, 2012), and even pragmatic abilities (van Compernelle & Kinginger, 2013). DA has shown great potential in providing diagnostic information regarding learners' emerging abilities as well as in promoting further development. However, despite being an essential part of language learning and assessment, vocabulary has largely evaded DA researchers' attention in L2 education. While some DA studies have investigated VK development with early-aged L1 learners or in L1 special education (e.g., Larsen & Nippold, 2007), empirical studies on DA of L2 vocabulary acquisition remain scarce.

Because of the complex nature of VK and the differences between L1 and L2 vocabulary acquisition, it is necessary to investigate the application of DA in L2 vocabulary acquisition separately. In an earlier literature review, Karami et al. (2019) examined four studies on DA of vocabulary to address two key questions: 1) how does DA, as an alternative to traditional assessment, meet the theoretical and practical criterion for assessing vocabulary? and 2) how effective is DA in ESL classroom practice for vocabulary assessment. Their review revealed that DA is effective in promoting vocabulary development and is beneficial for classroom

instruction. However, their review was limited to four studies and did not consider vocabulary assessment within the context of larger, more comprehensive language skills.

In this paper, we systematically investigate the existing literature on DA of vocabulary and the effect of DA procedures on adult learners' VK. As mentioned earlier, the body of literature specifically addressing DA of vocabulary is relatively small. As such, this review will also include studies of DA procedures related to L2 reading, writing, listening, and speaking with interventions targeting vocabulary. The literature review aims to answer the following research questions:

**RQ1:** What aspects of vocabulary knowledge have been targeted in the existing research, and what methods and instruments have been reported in the literature for vocabulary-focused DA?

**RQ2:** What mediation strategies have been used by DA researchers to help assess and promote vocabulary development? How were they designed?

**RQ3:** In what ways does DA offer additional insights into learners' vocabulary development and what types of information have been reported?

**RQ4:** How does the existing literature address the potential of DA to enhance L2 vocabulary acquisition and promote long-term retention of L2 vocabulary knowledge?

## **Method**

As suggested by Risko et al. (2008), we followed a systematic process for selecting and analyzing studies to answer these questions.

### *Selection Criteria*

We employed a two-step strategy to locate related studies. The first step involved a database search, during which we conducted two searches in ProQuest Linguistics and Language Behavior Abstracts (LLBA) and Educational Resources Information Center (ERIC). LLBA and ERIC are two of the most important databases in the field of L2 education (Lin, 2019). LLBA, as a part of the Social Science Premium collection, indexes literature in linguistics and language science from journal articles, books, book chapters, and dissertations from over 600 sources of publications (ProQuest LibGuides, n.d.). ERIC, recognized as the largest and most widely used education-related database in the world, contains millions of journal articles, research papers, conference papers, books, and dissertations (ProQuest LibGuides, n.d.). The combination of these two databases provides comprehensive coverage of the topic under discussion.

In both LLBA and ERIC, we searched “dynamic assessment” with “vocabulary” and “second language” as main subjects. For LLBA, we used the terms “vocabulary”, “vocabulary instruction”, “vocabulary learning”, and “vocabulary size” with the terms “English as a second language”, “English as a second language learning”, “English as a second language instruction”, “English as a second language teaching material”, “English as a second language teaching methods”, and “English as a second language tests”; for ERIC, we used the terms “vocabulary”, “vocabulary development” and “vocabulary skills” with the terms “English (second language)”, “second language instruction”, and “second language learning”. Our initial search returned 34 results in LLBA and 17 in ERIC. After removing duplications ( $n = 2$ ), the first search yielded a collection of 49 articles.

The second part of our search was guided by Poehner and Wang's (2021) timeline research article, which provided an overview of DA in the context of L2 learning and identified four major themes in L2 DA research. Given its comprehensive nature, this article served as an important resource for our investigation. Among the 74 studies they examined, three focused on vocabulary ( $n = 3$ ).

### *Exclusion Criteria*

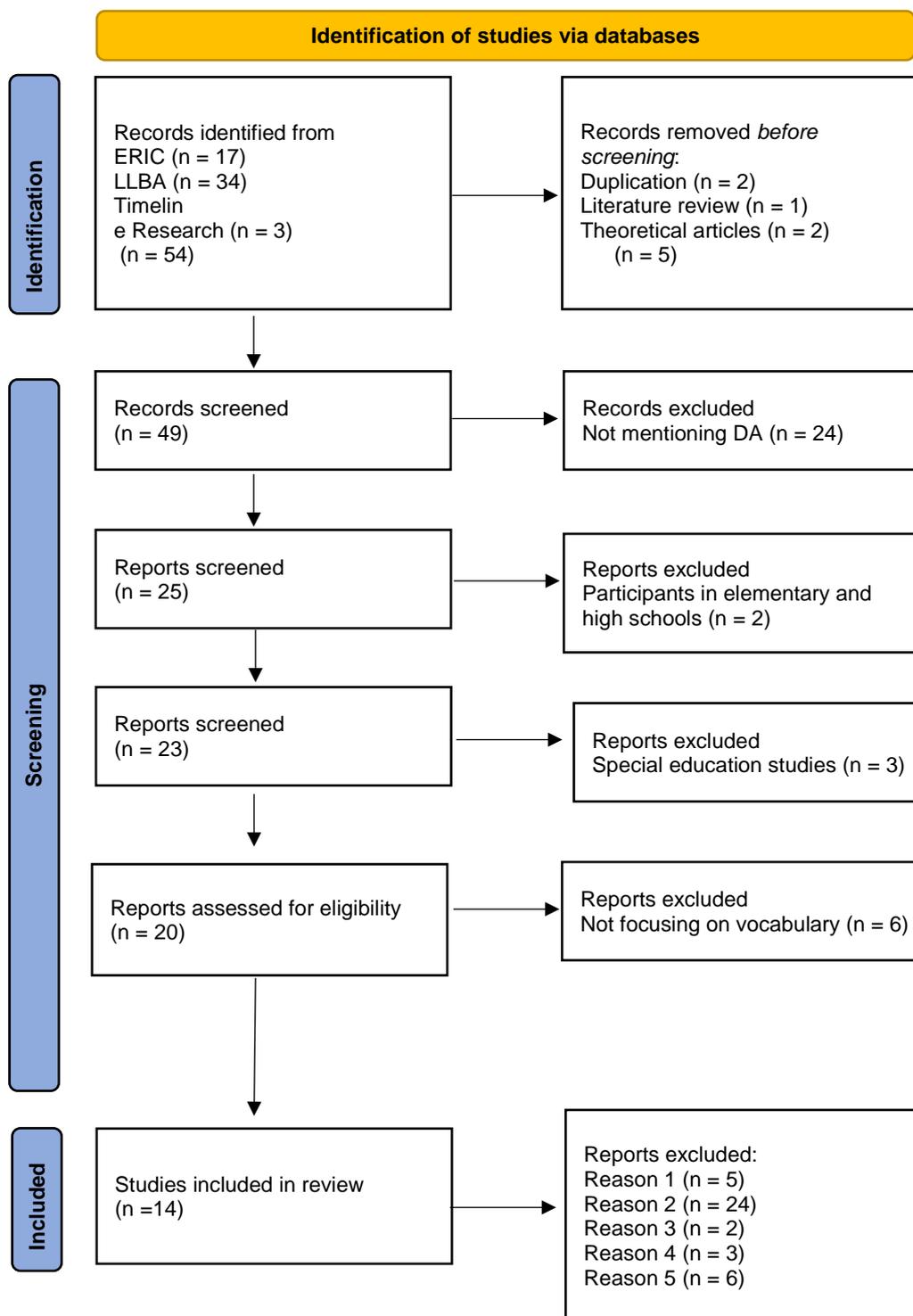
First of all, we aimed to focus on empirical studies that explored the effect of DA on vocabulary development; therefore, we excluded all literature reviews and theoretical discussion articles ( $n = 2$ ). Second, we prioritized studies that intentionally employed DA procedures to investigate the results and effects of DA, leading us to exclude articles that did not mention DA in their method sections, such as those that only addressed DA as a potential future research direction ( $n = 24$ ).

Additionally, while the DA procedure has been widely used in many different areas, our focus was specifically on its implementation in second and foreign language acquisition. Extensive research has demonstrated significant differences in the ways learners acquire vocabulary in L1, L2, and bilingual/multilingual learning contexts. Due to the lack of a clear distinction between bilingual/multilingual education and second/foreign language education, we excluded studies with participants in elementary and middle schools within an ESL context, which means studies with ESL participants younger than 14 years old were also excluded ( $n = 2$ ). We also excluded studies in the field of special education that dealt with learners with various learning difficulties ( $n = 3$ ) and studies that mentioned vocabulary but did not specifically focus on vocabulary assessment or intervention ( $n = 6$ ). Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart detailing this process.

### **Results**

Table 2 provides general information about the 14 studies included in this review. For each study, this includes the title, year of publication, author(s), participant information, research question(s), DA approach and model, length of the intervention, mediation media used, and the specific skills targeted.

**Figure 1**  
*The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)*



**Table 2***General Information of the 14 Focal Studies*

Title	Year	Author(s)	Participants number gender (age)	Research Question(s)	Approach - Model	Length of intervention	Mediation Media	Focal skill
Effects of Dynamic Assessment on College EFL Learners' Reading Skills	2012	Teo	2 F, 3 M	(1) What are the effects of DA on promoting Taiwanese college EFL students' reading skills? (2) Does DA help the participants realize their learning potential and if so, to what extent?	interventionist - sandwich	Six weeks	human - one on one	Reading
Dynamic Assessment and the Impact on English Language Learners' Reading Comprehension Performance	2012	Naeini & Duvall	8 F, 2M	(1) Can mediation in DA influence EFL university students' reading comprehension performance? (2) How might mediation in DA improve the EFL university students' reading comprehension performance?	interactionist - sandwich	Three 30- minute sessions	human - one on one	Reading
Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation	2015	Poehner et al.	68 (listening), 82 (reading)	-	interventionist - cake	One session	computerized	Reading and listening

Applicability of Peer-dynamic Assessment in Crowded Second Language Classes	2016	Khoshsima & Rezaee	15 F (15-17)	(1) How may informed peer-DA contribute to co/constructing reading comprehension gains? (2) How can informed peer-DA assist learners to co/building vocabulary knowledge?	interventionist - cake	Ten one-hour sessions	human - group	Reading
Dynamic Assessment of Word Derivational Knowledge: Tracing the development of a Learner	2016	Leontjev	1 (16)	How, if at all, does dynamic assessment promote L2 English learners' ability to derive words?	interventionist - sandwich	Two human-mediated and two computerized sessions	human and computerized	Two reading passages from textbook
Dynamic Assessment of Incidental Vocabularies: A Case of Iranian ESP Learners	2016	Hanifi et al.	25	Does dynamic assessment improve ESP learners' incidental vocabulary learning?	interventionist - sandwich	-	human - group	Lexical inference
Dynamic assessment of morphological awareness in the EFL context	2017	Hamavandi et al.	25 (14-18)	(1) What is the effect of dynamic assessment on EFL learners' reading comprehension? (2) To what extent can DATMA and Test of Morphological Structure (TMS) scores predict EFL learners' reading comprehension following the dynamic assessment treatment sessions?	interventionist - sandwich	-	human - group	Morphological awareness

ZPD-Based Dynamic Assessment and Collaborative L2 Vocabulary	2017	Mirzaei et al.	50 F (15-16)	(1) Does cumulative G-DA have any significant effect on enhancing long- and short-term gains in depth of vocabulary knowledge of Iranian EFL learners? (2) How can cumulative G-DA result in gains in the learners' depth of vocabulary knowledge in terms of transition from other-regulation to self-regulation?	interactionist-sandwich	Eight sessions	human - group	Meaning recognition
Exploring lexical inferencing as a vocabulary acquisition strategy through computerized dynamic assessment and static assessment	2018	Ebadi et al.	34 F, 38 M (16-24)	(1) more success in lexical inferencing? (2) better transfer of lexical inferencing skill to more difficult texts, and (3) better acquisition and retention of inferred words?	interventionist - sandwich	Five sessions	computerized	Lexical inference
Effect of interactionist dynamic assessment on English vocabulary learning: Cultural perspectives in focus	2019	Ghonsooly & Hassanzadeh	65 F 55M (18-30)	(1) Does interactionist dynamic assessment differently affect the vocabulary learning of male learners with different cultural dimensions? (2) Does interactionist dynamic assessment differently affect the vocabulary learning of female learners with different cultural dimensions?	interventionist - sandwich	Ten sessions	human - group	Meaning recognition

The Effects of Vocabulary Enhancement Exercises and Group Dynamic Assessment on Word Learning Through Lexical Inferencing	2019	Bahramlou & Esmaeili	26 F, 19 M (13-19)	Among VE exercises, GDA, and GDA plus VE exercises, which condition would lead to more word learning through lexical inferencing?	interventionist, sandwich	Six sessions	human - group	Lexical inference
Investigating the role of dynamic assessment on speaking accuracy and fluency of pre-intermediate EFL learners	2020	Maryam Safdari & Jalil Fathi	62 (18-30)	(1) With speaking accuracy scores as covariate, how does dynamic assessment affect the speaking accuracy of Iranian pre-intermediate EFL learners? (2) With speaking fluency scores as covariate, how does dynamic assessment affect the speaking fluency of Iranian pre-intermediate EFL learners?	interventionist - sandwich	8 interventions	human	Speaking
Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis	2023	Jeon	15 F, 20 M (12)	(1) Is CA-DA effective for receptive and productive vocabulary learning in L2 learners? (2) How can CA-DA be used for diagnosing L2 learners' vocabulary knowledge?	interventionist - sandwich	Two 25 minutes sessions	computerized-chatbot	Meaning recognition

---

<p>The effects of text-based and audio-based dynamic glosses on L2 vocabulary learning: a dynamic assessment approach</p>	<p>2023</p>	<p>Rassaei</p>	<p>57 (21-28)</p>	<p>(1) Is text-based dynamic glossing (TDG) effective for L2 vocabulary learning in terms of remembering the meaning of unfamiliar words?                  (2) Is audio-based dynamic glossing (ADG) effective for L2 vocabulary learning in terms of remembering the meaning of unfamiliar words?                  (3) Are TDG and ADG more effective than non-dynamic glossing (NDG)?                  (4) Is there a statistically significant difference between the effectiveness of TDG and ADG for L2 vocabulary learning?</p>	<p>interventionist, sandwich</p>	<p>Two sessions</p>	<p>Google Meet (text and audio)</p>	<p>Meaning recognition</p>
---	-------------	----------------	-------------------	---	----------------------------------	---------------------	-------------------------------------	----------------------------

---

*RQ1: What aspects of vocabulary knowledge have been targeted in the existing research, and what methods, and instruments have been reported in the literature for vocabulary-focused DA?*

Among the 14 articles reviewed, 10 focused specifically on vocabulary. Two articles focused on morphological awareness, an important aspect of VK. Nation (2021) mentioned that morphological awareness can help improve L2 vocabulary acquisition because it increases the quantity and quality of repetition. Leontjev (2016)\*<sup>2</sup> used both human-mediated and computerized DA to investigate a learner's derivational knowledge, following the interventionist sandwich model. Without specification, it seems that the researcher had developed his material for assessing derivational knowledge based on Bauer and Nation's (1993) affix levels. Although target words were presented in a sentence, no other contextual information was provided. During the human-mediated DA sessions, learners were asked to complete classification exercises and affix elicitation exercises; the C-DA sessions used multiple-choice tasks. Hamavandi et al. (2017)\* also focused on morphological awareness. Adapting the Dynamic Assessment Task of Morphological Assessment (DATMA) (Larsen & Nippold, 2007) with target words selected from the Educator's Word Frequency Guide (Zeno et al., 1995), the authors aimed to determine whether participants could use their knowledge of familiar base words and suffixes to determine the meanings of uncommon morphologically complex words. Their research followed the interventionist approach, using the Nelson-Denny reading test as the pre-test, and the test of morphological structure (Carlisle, 2000) as the post-test.

Four studies took an interest in lexical inferencing. According to Nation (2021), lexical inferencing, or the ability to infer the meaning of an unfamiliar word, is crucial in L2 vocabulary acquisition. Hanifi et al. (2016)\* used reading texts from the textbook adopted by their participants' university. Although the authors did not specify their approaches, judging from their description and sample items, they likely followed the interactionist sandwich design, with two vocabulary tests as pre- and post-tests. Their group DA employed multiple-choice questions. Ebadi et al. (2018)\* also focused on lexical inferencing. The authors developed a computerized DA (C-DA) program to deliver mediation. Before the intervention, the participants were given five checklists as the pre-test. The C-DA intervention included five carefully chosen texts with four target words in each text and was administered over five weeks. The five texts had an average of 337 words and less than 3% of unfamiliar words. After reading each text, learners were presented with reading comprehension questions in static assessment. Following the same design, near and far transfer texts were administered. The control group used the same text and question format. These target words were tested again one day and four weeks later for long-term retention. In addition to the study included here, the researchers also developed a Persian version of the C-DA intervention, although a discussion of that version is out of the scope of this review.

Bahramlou and Esmaeili (2019)\* investigated the effect of group DA (G-DA) on lexical inferencing by comparing the results between the use of G-DA, vocabulary exercises (VE), and their combination (G-DA+VE). They used six texts with an average length of 1,042 words, in

---

<sup>2</sup> We have designated all focal articles included in our review with \* to distinguish them more clearly from other studies cited in the paper.

which 94.14%-96.93% of the words were among the first to third 1,000 most frequent words, with target words highlighted in the texts. The Vocabulary Levels Test (Nation & Beglar, 2007) was used to determine learners' vocabulary levels, and the Vocabulary Knowledge Scales (VKS) (Wesche & Paribakht, 1996) were adopted as their pre- and post-test. Khoshsima and Rezaee (2016)\* used 10 passages from Anderson's book *Active Skills for Reading* as the input for their reading comprehension test. Their intervention consisted of 10 one-hour G-DA sessions. In each session, the learners first read a passage with embedded target vocabulary and then teamed up to understand the main topic and the meaning of the target words collaboratively following a brief introduction from the teacher. Based on the episodes they provided, their vocabulary mediation focused on lexical inference.

Four studies investigated the effect of DA on meaning recognition. Focusing on vocabulary learning, Ghonsooly and Hassanzadeh (2019)\* took a very different approach to investigate the effect of DA on learners with different cultural perspectives. The researchers selected 40 unknown words (learners rated as "I do not remember having seen this word before" or "I have seen this word before, but I do not know what it means") from *Vocabulary in Use*, a textbook for vocabulary development with vocabulary and practice exercises side by side. They used the VKS as their pre-test and developed their own instrument for the post-test. The cultural dimensions questionnaire (Saboori et al., 2015) was used to identify learners' cultural perspectives. According to Breen (2001), gender and power relationships are important parts of culture and are crucial to interaction. The authors looked into the notions of collectivism and individualism specifically in their literature review section, hypothesizing that learners with collectivist culture might benefit less from DA, as they were expected to be recipients of knowledge, while learners with individualism orientation might benefit more due to their willingness and readiness to explore and engage. As an innovative attempt, Jeon (2023)\* utilized a chatbot to explore the effect of chatbot-assisted DA on vocabulary learning. Built with Google's Dialog flow, the text-based chatbot allowed learners to interact using both Korean and English and tested learners on both receptive and productive vocabulary knowledge. Her participants were divided into three groups, the chatbot-assisted DA group (CA-DA), the chatbot-assisted non-DA group (CA-NDA), and a control group. Participants were asked to read two sections of modified reading materials (209 words and 221 words) with 10 target words during the two treatment sessions. Using the VKS, Mirzaei, et al. (2017)\* conducted eight G-DA sessions with 25 female learners. In each session, learners were asked to translate sentences into English, following the interactionist cumulative group DA approach. Rassaei's (2023)\* dynamic glossing (DG) intervention focused on 14 target words that were chosen based on learners' performance in the pre-test. These target words were embedded in six reading passages. During the two treatment sessions, learners interacted with mediators via text or audio to figure out the meanings of the target words. The author also developed a C-DA to measure learners' vocabulary development. Target words were also embedded in two reading passages, and learners were required to identify the correct L1 translations.

Three Studies focused on reading or listening comprehension. Teo (2012)\* used the TOEFL sample tests to assess learners' reading skills, including the use of contextual clues in predicting word meanings. Adopting the sandwich model, Teo provided one mediation session per week for four weeks. The mediator used two self-designed flowcharts to guide the mediation, but it was unclear whether the interventionist or interactionist approach was

followed. Naeini and Duvall (2012)\* used reading comprehension tests for both pre- and post-tests, among which 13 out of 29 questions focused on word meaning. Each mediation session lasted 30 minutes and targeted main ideas, inferences, or word meaning, following the interactionist sandwich format to provide individualized support for learners. Designed to investigate Chinese reading and listening comprehension, Poehner et al. (2015)\* designed a C-DA to integrate interventionist DA in the cake format. Adapted from the Chinese Proficiency Test (HSK), their reading comprehension test consisted of 24 items targeting vocabulary, structure, discourse, and genre, and their listening test consisted of 23 items targeting phonetics, vocabulary, grammar, pragmatics, contextual inference, and information-seeking.

Finally, focusing on speaking accuracy and fluency, Safdari and Fathi (2020)\* provided mediation intervention on vocabulary and grammar. The Preliminary English Test (PET) speaking part was used as both a pre- and post-test. Their vocabulary mediation used exercises from *English Vocabulary in Use* (Redman, 2003), following the interventionist sandwich model.

*RQ 2: What mediation strategies have been used by DA researchers in the field to help assess and promote vocabulary development? How were they designed?*

Three studies conducted one-on-one in-person mediation. Naeini and Duvall (2012)\* conducted in-person one-on-one mediation sessions. To plan their mediation, the authors conducted interviews with participants to determine their difficulties. Their mediation included a wide range of strategies, such as questions, hints, and explanations. Their study appeared to follow an interactionist approach, although they did not provide any information regarding the order of their mediation. Relying on Larsen and Nippold's (2007) graduated prompting system, Hamavandi et al (2017)\* conducted one-on-one mediational sessions with their participants. Six mediational moves were planned and arranged on the implicit to explicit continuum, such as pauses, leading questions, hints, and explanations. Teo (2012)\* developed a flowchart to guide mediation. The intervention was provided once a week for four weeks, in which mediation was also provided from implicit to explicit, with the amount of mediation based on the number of incorrect answers.

Among all the studies reviewed, three used computers to deliver mediation. As one of the first attempts to computerize DA, Poehner et al. (2015)\* added one distractor to the original four-option multiple-choice questions to allow learners multiple chances for mediation. Mediation was arranged from implicit to explicit. They also provided a sample of their mediational moves. Ebadi et al (2018)\* also designed a computerized system to deliver mediation. Following a similar design, their mediation focused on gradually highlighting the more relevant portion.

Using Dialogflow, Jeon (2023)\* designed an AI-assisted chatbot to deliver mediation. The author drew on Poehner et al.'s (2015)\* C-DA study and Nation's (1990) five-step procedure to design her initial mediation. The author also conducted a pilot study to develop the initial version of the prompts and adjusted them according to learners' responses. Samples of mediational moves were provided.

Two studies used both human and computerized mediation. Leontjev (2016)\* built on his previous study in planning mediation. Judging from the samples provided from both human and computerized DA, he provided mediation from the most implicit to the most explicit.

Rassaei's (2023)\* treatment included two parts. The first part used dynamic glossing in both text and audio form, during which the mediator provided learners with graduated responses to help them identify the meanings of unknown words. Mediation prompts included four moves from encouraging learners to reread the sentence to providing the target words in a more transparent context. The second part was C-DA intervention, in which the same set of prompts were used.

In response to the efficiency of DA, six studies utilize group DA. In Hanifi et al. (2016)\*, mediation was presented to students in a group. They did not provide specific information regarding the mediation, but their example suggested that mediation included three leveled hints and was arranged from implicit to explicit. Similarly, Mirzaei et al (2017)\* did not provide detailed information regarding their mediation design, but their samples suggested that the authors also used graduated prompts from implicit to explicit. Interestingly, the authors mentioned that during group DA, learners were given opportunities to conduct private speech as a mediational tool. Ghonsooly and Hassanzadeh (2019)\* also designed group mediation. Seven stages of mediation, ranging from guessing to paired definition matching, focused on morphological awareness, synonyms and antonyms, and explicit definition. Safdari and Fathi (2020)\* delivered mediation in groups as follows: the mediator asked learners to provide explanations of their choices and, in the case of a wrong choice, directed learners' attention to the context in which the target word was used. Bahramlou and Esmaili (2019)\* used a slightly different group DA design. If the learners answered a question correctly, they were asked to provide explanations. If they provided an incorrect answer, the mediator elicited answers from a different learner before providing more explicit mediation. Khoshshima and Rezaee (2016)\* employed a peer group DA design, asking their learners to employ five levels of mediation for peer DA following three training sessions.

*RQ3: In what ways does DA offer additional insights into learners' vocabulary development, and what types of information have been reported?*

Among all reviewed studies, only four mentioned the differential effect of DA on providing more fine-tuned information regarding learners' VK. Teo (2012)\* discussed whether DA could help participants realize their learning potential. To answer this question, the author calculated learning potential scores (LPS) based on Kozulin and Garb's (2002) formula. Her results showed that students' LPSs varied from 0.58 to 1.21. In addition, the author analyzed learners' LPSs in relation to their pre- and post-test scores and to each individual construct and found that DA provided more fine-tuned information regarding learners' development.

Poehner et al. (2015)\* calculated participants' actual scores and mediated scores and reported significant differences between them (listening:  $t = 16.055$ ,  $p < .001$ , Cohen's  $d = 1.020$ ; reading:  $t = 20.750$ ,  $p < .001$ , Cohen's  $d = 1.242$ ). These differences indicated that DA could provide more information regarding learners' development. More interestingly, they compared learners with the same actual score, mediated score, and LPS and found that their responses to mediation varied by construct. In the case of vocabulary, these learners required different amounts of mediation, indicating that they were distinctive in terms of vocabulary development.

Ebadi et al. (2018)\* compared the C-DA group's mediated and unmediated lexical inferencing scores and found that learners with similar unmediated scores varied widely in their

mediation score, confirming C-DA's differentiating ability. The authors also calculated LPSs and reported a moderate correlation between the learners' LPSs and their scores on transfer tasks ( $r = .37, p < .01$ ), indicating that the LPS could provide useful information about the learners' learning potential. They also highlighted that the information not only helped them understand learners' different levels of development for better placement but was also important for planning future instructional intervention.

Similarly, Jeon (2023)\* also calculated each learner's actual score and mediated score for the number of words for which they provided the correct definition in Korean or English. Interaction records between the learners and the chatbot were generated and analyzed. For receptive vocabulary knowledge, the one-way ANOVA results indicated that the three groups (CA-DA, CA-NDA, and control) performed significantly differently; Scheffe's post hoc analysis showed that the CA-DA group significantly outperformed the other two groups ( $p < .05$ ) in both the post-test and delayed post-test. The results showed that learners' unmediated and mediated scores both improved in the CA-DA group. The interaction data further showed that learners with the same unmediated scores exhibited different levels of development.

Four studies briefly addressed the diagnostic function of DA but did not provide detailed analyses or interpretations. Leontjev (2016)\* briefly addressed the diagnostic aspect of DA through his analysis of mediation. He found that the amount of mediation required by the learner decreased over time, even though the learner was still not yet at the stage of self-regulation. Khoshsima and Rezaee's (2016)\* qualitative analysis of mediation and learners' responses led them to conclude that DA could help diagnose learners' difficulties by providing the "reasons behind language learners' poor functioning" (p.934). From a cultural perspective, Ghonsooly and Hassanzadeh (2019)\* analyzed learners' vocabulary scores in relation to their genders and different cultural dimensions. Their results revealed significantly different effects of DA on learners with different cultural orientations along the collectivism/individual dimensions. Rassaei (2023)\* calculated the total actual and mediated scores. The unmediated scores indicated that learners in all groups performed statistically differently in the post-DA sessions. A more detailed analysis shows that learners' mediated scores improved in the second DA session, especially for the two dynamic glossing groups., indicating that these learners were able to "identify correct definitions for the vocabulary items from more implicit mediation," which the author interpreted as evidence of their "vocabulary development" (p.517).

The remaining six studies did not address the assessment aspect of DA in terms of providing information regarding learners' emerging abilities. Naeini and Duvall (2012)\* reported both quantitative data and interaction data. Their quantitative data did not show any detailed diagnosis, but their interactive data showed that learners demonstrated different difficulties regarding reading strategies. Hanifi et al. (2016)\*, Hamavandi et al. (2017)\*, Mirzaei et al. (2017)\*, Bahramlou and Esmaeili (2019)\*, and Safdari and Fathi (2020)\* did not touch upon assessing emerging abilities.

*RQ4: How does the existing literature address the potential of DA for enhancing second language vocabulary acquisition and promoting long-term retention of vocabulary knowledge?*

All but one of the reviewed studies mentioned that DA helped promote vocabulary acquisition. With a focus on reading and listening, Poehner et al (2015)\* did not address the effect of DA on learners' vocabulary development. All other 13 studies reported data for vocabulary learning/development.

Teo (2012)\* analyzed the pre- and post-test score differences of five participants using a paired sample t-test, revealing a significant improvement in reading scores ( $t(5) = -4.118, p = 0.015$ ). This led Teo to conclude that DA helped improve learners' reading skills. He also conducted a quantitative analysis of the mediation and found that DA helped improve the learners' motivation to use English. Naeini and Duvall (2012)\* reported learners' pre- and post-test raw scores in an appendix without discussing them or performing any statistically analysis, although these scores suggested learner improvement following mediation for items related to word meanings.

Khoshsima and Rezaee (2016)\* did not provide details on the structure of their mediation, but they observed that DA helped learners "master" word meanings in a "more useful way" (p.934). Leontjev (2016)\* compared learner performance across four static assessment (SA) sessions. The author found that learner performance improved most between the first two SA sessions, indicating that human-mediated DA was effective. Learner performance also improved to a smaller extent between the third and fourth SA sessions, indicating that C-DA also helped improve learner performance. The author attributed the smaller improvement to the ceiling effect. The learner demonstrated the ability to use new affixes not covered in any DA session, indicating a transfer effect. More interestingly, using think-aloud protocols and interviews, the author found that the learner's knowledge source shifted from syntactic to semantic knowledge and analyzed words more frequently with more varied strategies. The author concluded from the qualitative data that DA helped draw the learner's attention to analyzing the words without explicit instruction. Hanifi et al (2016)\* found that learners' mean post-test scores were significantly higher than their pre-test scores ( $t(24) = 13.936, p < .001$ ), indicating that DA was effective in promoting vocabulary learning. Ebadi et al. (2018a)\* calculated learners' mean scores on lexical inferencing, near transfer, far transfer, acquisition, and retention. A one-way MANOVA revealed a significant difference between the DA group and the control group ( $F(6,65) = 27.573, p < .001$ ). Post-hoc analyses showed that, except for far transfer, the DA group outperformed the control group in all other aspects, indicating that DA effectively improved learners' lexical inferencing.

Also focused on morphological awareness, Hamavandi et al. (2017)\* reported that the DA group scored significantly higher than the non-DA group ( $t(48) = 3.13, p < .001$ ), indicating that DA of morphological awareness improved learners' reading comprehension. Mirzaei et al. (2017)\* performed a split-plot ANOVA for the two groups across the pre-test and two post-tests, the results of which revealed a significant interaction between time and group (Wilks'  $L = .497; F(2, 47) = 23.807, p < .001, \eta^2_p = .503$ ), indicating that group DA was considerably more effective in helping learners develop the depth of L2 vocabulary knowledge than the non-DA approach. Although the results of their delayed post-test showed that both groups

experienced a decrease in vocabulary knowledge, the DA group retained considerably more knowledge than the control group.

Ebadi et al. (2018)\* compared learners' performance in C-DA and SA groups, and their results showed that the assessment framework significantly affected learners' performance ( $F(6, 65) = 27.573, p < .001$ ; Wilk's  $L = 0.282$ ). They also found that C-DA led to significantly better performance on lexical inferencing, near transfer, and acquisition and retention tasks, but not on far transfer tasks ( $F(6, 65) = 3.279, p = .074$ ). They concluded that, although with regression and backsliding, the DA framework enabled learners to transfer what they had learned to more complex tasks.

Ghonsooly and Hassanzadeh (2019)\* compared the scores between learners with different cultural orientations and found that DA is most effective for learners with an individualist/collectivist orientation. They did not provide any detailed information regarding vocabulary learning specifically but concluded that the individualism/collectivism orientation was associated with DA vocabulary learning. Bahramlou and Esmaeili (2019)\* reported learners' post-test and delayed post-test scores as evidence of learning and retention. Their one-way MANOVA analysis showed that for the depth of vocabulary knowledge, group conditions significantly affected learning ( $F(2, 42) = 4.70, p < 0.014^3, \eta^2_p = .18$ ) and retention ( $F(2, 42) = 7.22, p < 0.002, \eta^2_p = .25$ ); and for the size of vocabulary knowledge, group conditions also significantly affected learning ( $F(2, 42) = 4.06, p < 0.024, \eta^2_p = .16$ ) and retention ( $F(2, 42) = 13.18, p < 0.000, \eta^2_p = .386$ ). Specifically, while the difference between the G-DA and VE groups was not significant, the G-DA+VE group performed significantly better than the other two groups. These results indicated that DA was effective in improving the depth of learners' vocabulary knowledge. In addition, in terms of the size, the G-DA+VE group significantly outperformed the G-DA and VE groups on both learning and retention, indicating the effectiveness of combining G-DA with vocabulary exercises. Safdari and Fathi (2020)\* performed ANCOVA and their results showed that DA significantly improved speaking accuracy ( $F = 267.58, p = 0.001$ ) but not fluency ( $F = 0.92, p = 0.34$ ). They also conducted five interviews with participants, who indicated that DA could help discover and correct their problems, leading to accuracy improvements, but it did not help improve their fluency.

Jeon (2023)\* reported that for receptive vocabulary knowledge, the three groups performed significantly differently in post- ( $F(2, 50) = 59.696, p < .05, \eta^2 = 0.70$ ) and delayed post-test ( $F(2, 50) = 59.696, p < .05, \eta^2 = 0.70$ ); and for productive vocabulary knowledge, the three groups performed significantly differently in post- ( $F(2, 50) = 21.640, p < .05, \eta^2 = 0.46$ ) and delayed post-test ( $F(2, 50) = 16.277, p < .05, \eta^2 = 0.39$ ). Specifically, the CA-DA group significantly outperformed the CA-NDA group and the control group in both the post-test and delayed post-test ( $p < .05$ )<sup>4</sup>. Qualitative analysis also showed that learners in the CA-DA group gradually required less mediation in terms of both the amount and the degree of explicitness. The author attributed the superior performance of the CA-DA group to ZPD-based mediation (p.1356) and reduced cognitive load (p.1357). In Rassaei's (2023)\* analysis of DA scores, the two dynamic glossing groups outperformed the non-dynamic and control groups, indicating that dynamic glossing, both in text and audio, helped learners develop vocabulary knowledge.

<sup>3</sup> For studies that did not report specific  $p$  values or degrees of freedom, we included statistical information as reported in the original studies.

<sup>4</sup> Only  $p$  values were reported of all post hoc comparisons in the original study.

The results also show that learners in the dynamic glossing group obtained significantly higher mediated scores compared than those in the non-dynamic group in the second DA session, although the differences were not significant in the first DA session. The author thus concluded that text and audio-based dynamic glossing of L2 new words was more effective than the conventional non-dynamic method. More interestingly, the author also found that the text-based dynamic glossing group outperformed the audio-based dynamic glossing group in the second DA session, suggesting that text-based dynamic glossing was more effective than the audio-based for vocabulary learning. The author noted that traditional non-dynamic vocabulary tests could not capture such a difference.

## **Discussion**

The purpose of this review was to systematically examine the current literature on Dynamic Assessment (DA) of vocabulary. Overall, our review indicates that L2 DA research has not yet paid enough attention to vocabulary knowledge. To answer the first research question, of the 14 studies that focused on vocabulary acquisition, six specifically focused on such aspects as lexical inferencing and morphological awareness, while others focused primarily on meaning recall/recognition. According to Nation (2021), a well-balanced L2 vocabulary instruction should include deliberate teaching of vocabulary as well as teaching learners to take responsibility for vocabulary learning. Including vocabulary acquisition strategies can promote autonomy and push vocabulary knowledge development in the long term. Peña et al. (2001) also mentioned that mediation could be designed to target the principle of task solution or problem-solving strategies. However, a closer look at the studies reviewed revealed that the complexity of vocabulary knowledge was not rigorously mediated in them. For example, for lexical inferencing, the mediation usually involved indicating an error, highlighting the part indicative of the meaning, and explicitly explaining the word. These steps did not really help learners identify the relevant part on their own, which is critical for achieving self-regulation. In addition, other vocabulary learning strategies, such as the use of dictionaries, have not been addressed in any of these studies. Further, all these studies focused on a single aspect of vocabulary knowledge, obscuring the complexity of vocabulary knowledge.

The findings also indicated a clear preference among DA researchers for the interventionist model. All but one of the reviewed studies adopting this model, among which 11 used the sandwich format, and two used the cake format. Only one study utilized the interactionist-sandwich model. The predominance of the interventionist model in these studies may reflect practical considerations within DA research. According to Poehner and Lantolf (2013), although the interactionist approach aligns more closely with Vygotskian principles and is considered ideal, it is also more challenging and demanding. The interventionist model, on the other hand, is more standardized. Its pre-structured mediation provides more readily comparable results and places less pressure on the mediator, making DA more accessible and more broadly applicable, including in a computerized form. Additionally, the sandwich format, also known as the pretest-mediation-posttest model, makes it easier for researchers to compare learners' development before and after the intervention to assess the effects of mediation.

In the studies reviewed, DA researchers utilized a wide variety of instruments, such as TOEFL, Chinese Proficiency Test, Preliminary English Test, Vocabulary Levels Test, Vocabulary Knowledge Scales, and Vocabulary Checklist. One study adapted the Dynamic

Assessment Task of Morphological Assessment. The variety of instruments used by DA researchers reflects the flexibility and adaptability of DA in different learning contexts. Researchers also utilized textbooks and their accompanying vocabulary exercises in their study design. This integration shows that DA can be used to adapt more traditional classroom materials to unite assessment and instruction (Poehner & Lantolf, 2005). The wide application of DA shows that, as Poehner (2008) maintains, DA can be applied to any instrument as a method, pointing to its feasibility in different educational contexts. More interestingly, among the 14 focal studies, five developed their own instruments to use either as standalone tools or with other existing instruments. This trend suggests that many DA researchers found that existing assessment instruments are insufficient.

With regard to our second research question, our review revealed that existing DA vocabulary research has insufficiently focused on the development of mediation strategies for vocabulary. Mediation is the most important element of DA that distinguishes it from non-DA forms of assessment. However, several studies did not provide any information on the mediation used in the DA procedure. Most studies adopted the mediation strategies developed for reading or writing and followed Aljaafreh and Lantolf (1994) in arranging mediation from the most implicit to the most explicit, but the rationale for using such mediation strategies for vocabulary was rarely explicitly discussed. L2 research has shown that the process of vocabulary acquisition is not exactly the same as that of the acquisition of reading or writing (Nation, 2021). As such, it is necessary to carefully evaluate whether the mediation strategies developed for reading or writing research can be directly used for vocabulary. More importantly, future research on DA of vocabulary should more systematically explore how to optimally mediate L2 vocabulary development in DA procedures. The design of DA vocabulary mediation should consider how to take into account the complex nature of vocabulary knowledge (Nagy & Scott, 2000). It should also take full advantage of findings from vocabulary instruction research to integrate aspects such as meaning-focused output, vocabulary learning strategies, and fluency development (Nation, 2021).

To answer our third research question, our review shows that the field has an insufficient understanding of DA's ability to provide additional information regarding learners' vocabulary knowledge. Despite the fact that DA's theoretical foundation allows for assessing both independent and emerging abilities, only four of the 14 focal studies elaborated on DA's ability to provide information regarding learners' emerging vocabulary knowledge by comparing learners' independent performance and mediated performance, and four others briefly mentioned the diagnostic aspect. This limited focus suggests that the existing research has not fully leveraged DA to provide additional information.

The most important information revealed by DA design is that learners with the same or similar independent performance may have very different mediated scores, indicating that vocabulary development varies widely among learners, even if they initially appear to be at similar levels. Among the additional information provided by DA, the LPS (Kozulin & Garb, 2002) has been mentioned in three studies. This score serves to quantify the ZPD, offering a measure of each learner's capacity for growth when provided with mediation. In addition, DA also provided information regarding the focus of future instruction. By pointing out the areas learners struggle independently but succeed with mediation, DA can provide valuable guidance for instructors in tailoring to learners' specific needs.

Our review underscores the positive impact of DA on vocabulary knowledge development. The findings show that DA is a powerful intervention for vocabulary acquisition, contributing to both immediate gains and long-term retention. However, as most studies employed the interventionist-sandwich approach, the analyses were predominantly quantitative. While this quantitative focus allows for clearer measurement, it may overlook some of the more nuanced insights that could emerge from a qualitative perspective. Given that DA assumes each individual's development is unique, relying solely on quantitative data may not fully capture the complexity of this process.

### **Conclusion**

The results of the current literature review show that studies of DA of vocabulary for L2 learners remain limited. The 14 reviewed studies focused on various aspects of vocabulary knowledge, including vocabulary acquisition, morphological awareness, lexical inferencing, and meaning recognition. These studies utilized a range of instruments, such as the Vocabulary Knowledge Scales, the Vocabulary Levels Test, and customized assessments. With a clear preference for the interactionist-sandwich model approach, the studies delivered mediation in diverse ways, ranging from one-on-one in-person sessions to computerized mediation. The mediation strategies adopted by DA researchers were usually arranged from implicit to explicit, although the rationale for doing so was often not explicitly discussed. Most studies reported that DA positively impacted vocabulary acquisition and retention, but only four discussed DA's ability to offer detailed insights into learners' vocabulary knowledge beyond standard assessments.

This review has identified several issues related to DA of vocabulary. First, all the studies reviewed focused on only one aspect of vocabulary knowledge. Given that vocabulary knowledge is a complex construct, focusing on only one aspect is insufficient to provide a comprehensive picture of learners' development (González-fernández & Schmitt, 2020) which DA is designed to capture. Future studies could address this limitation by developing multi-aspect, cross-sectional DA instruments that assess various facets of vocabulary knowledge. Second, most current DA studies utilized the interventionist-sandwich approach and thus produced mostly quantitative data on learners' vocabulary development and intervention outcomes. Future studies could benefit from adopting an interactionist approach to investigate the use of DA for vocabulary development through qualitative data, such as learners' reflections and observations of mediation sessions. Such data may help researchers capture a wider variety of information, including that pertaining to learner motivation and engagement, and gain a deeper understanding of vocabulary knowledge development at an individual level. Finally, this review revealed that the mediation used in existing studies was often insufficiently tailored for vocabulary knowledge, with many strategies adapted directly from reading and writing research. Future studies should draw on vocabulary assessment and instructional research to design and plan mediation specifically aligned with vocabulary development needs.

### **ORCID**

 <https://orcid.org/0009-0007-5960-3948>

 <https://orcid.org/0009-0002-9751-4376>

### Acknowledgements

Not applicable.

### Funding

Not applicable.

### Ethics Declarations

### Competing Interests

No, there are no conflicting interests.

### Rights and Permissions

### Open Access

This article is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which grants permission to use, share, adapt, distribute and reproduce in any medium or format provided that proper credit is given to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if any changes were made.

### References

- Ableeva, R. (2018). Understanding learner L2 development through reciprocity. In J. P. Lantolf, M. E. Poehner, & M. Swain (Eds.) *The Routledge handbook of sociocultural theory and second language development* (pp. 266–281). Routledge.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum.
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465–483. <https://doi.org/10.2307/328585>
- \*Bahramlou, K., & Esmaeili, A. (2019). The effects of vocabulary enhancement exercises and group dynamic assessment on word learning through lexical inferencing. *Journal of Psycholinguistic Research*, 48(4), 889–901. <https://doi.org/10.1007/s10936-019-09638-x>
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Breen, M. P. (2001). Navigating the discourse: On what is learned in the language classroom. In C. N. Candlin & N. Mercer (Eds.), *English language teaching in its social context: A reader* (pp. 306–322). Routledge.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66. <https://doi.org/10.1080/14703290601081332>
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12, 169–190. <https://doi.org/10.1023/a:1008131926604>
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38–63. <https://doi.org/10.125/44117>
- Davin, K. J., Troyan, F. J., & Hellmann, A. L. (2014). Classroom dynamic assessment of reading comprehension with second language learners. *Language and Sociocultural Theory*, 1(1), 1–23. <https://doi.org/10.1558/lst.v1i1.1>
- Donato, R., & McCormick, D. (1994). A sociocultural perspective on language learning strategies: The role of mediation. *Modern Language Journal*, 78(4), 453–464. <https://doi.org/10.2307/328584>
- \*Ebadi, S., Weisi, H., Monkaresi, H., & Bahramlou, K. (2018). Exploring lexical inferencing as a vocabulary acquisition strategy through computerized dynamic assessment and static assessment. *Computer Assisted Language Learning*, 31(7), 790–817. <https://doi.org/10.1080/09588221.2018.1451344>
- Feuerstein, R., Rand, Y. A., Hoffman, M., & Miller, R. (1979). Cognitive modifiability in retarded adolescents: effects of instrumental enrichment. *American Journal of Mental Deficiency*, 83(6), 539–550. <https://doi.org/10.1080/13638490310001655140>
- Feuerstein, R., Rand, Y., & Rynders, J. E. (1988). *Don't accept me as I am: Helping "retarded" people to excel*. Springer. <https://doi.org/10.1007/978-1-4899-6128-0>
- \*Ghonsooly, B., & Hassanzadeh, T. (2019). Effect of interactionist dynamic assessment on English vocabulary learning: Cultural perspectives in focus. *Issues in Educational Research*, 29(1), 70–88. <https://search.informit.org/doi/10.3316/ielapa.171745800903416>

- González-fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, 41(4), 481–505. <https://doi.org/10.1093/applin/amy057>
- Haastrup, K. (2008). Lexical inferencing procedures in two languages. In D. Albrechtsen, K. Haastrup, & B. Henriksen, *Vocabulary and writing in a first and second language* (pp. 67–111). Palgrave Macmillan. [https://doi.org/10.1057/9780230593404\\_3](https://doi.org/10.1057/9780230593404_3)
- \*Hamavandi, M., Rezai, M. J., Mazdayasna, G., & Hui, S. K. F. (2017). Dynamic assessment of morphological awareness in the EFL context. *Cogent Education*, 4(1), 1324254. <https://doi.org/10.1080/2331186X.2017.1324254>
- \*Hanifi, S., Nasiri, M., & Aliasin, H. (2016). Dynamic assessment of incidental vocabularies: A case of Iranian ESP learners. *Advances in Language and Literary Studies*, 7(2), 163–170. <http://dx.doi.org/10.7575/aiac.all.v.7n.2p.163>
- Haywood, H. C., & Lidz, C. S. (2006). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511607516>
- Hill, K., & Sabet, M. (2009). Dynamic speaking assessments. *TESOL Quarterly*, 43(3), 537–545. <http://www.jstor.org/stable/27785036>
- \*Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, 36(7), 1338–1364. <https://doi.org/10.1080/09588221.2021.1987272>
- Karami, A., Howlett, K. M., & Bowles, F. A. (2019). Implementing dynamic assessment in assessing vocabulary development: What the literature says. *Journal on English Language Teaching*, 9(2), 52–59. <https://doi.org/10.26634/jelt.9.2.15518>
- \*Khoshshima, H., & Rezaee, A. (2016). Applicability of peer-dynamic assessment in crowded second language classes. *Journal of Language Teaching and Research*, 7(5), 929–935. <http://dx.doi.org/10.17507/jltr.0705.13>
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, 23(1), 112–127. <https://doi.org/10.1177/0143034302023001733>
- Lantolf, J. P. (2000). Introducing sociocultural theory. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 1–26). Oxford University Press.
- Larsen, J. A., & Nippold, M. A. (2007). Morphological analysis in school-age children: Dynamic assessment of a word learning strategy. *Language, Speech, and Hearing Services in Schools*, 38(3), 201–211. [https://doi.org/10.1044/0161-1461\(2007\)021](https://doi.org/10.1044/0161-1461(2007)021)
- \*Leontjev, D. (2016). Dynamic assessment of word derivational knowledge: Tracing the development of a learner. *Eesti Rakenduslingvistika Ühingu Aastaraamat [Estonian Papers in Applied Linguistics]*, 12, 141–160. <https://doi.org/10.5128/ERYa12.09>
- Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. Guilford Press.
- Lin, J. (2019). Factors related to EFL/ESL learners' reading strategy use: A literature review. *Chinese Journal of Applied Linguistics*, 42(1), 38–52. <https://doi.org/10.1515/CJAL-2019-0006>
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351–372. <https://doi.org/10.1177/026553220101800403>
- Meara, P. (1997). Towards a new approach to modeling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 109–121). Cambridge University Press.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- \*Mirzaei, A., Shakibei, L., & Jafarpour, A. A. (2017). ZPD-based dynamic assessment and collaborative L2 vocabulary learning. *The Journal of Asia TEFL*, 14(1), 114–129. <https://doi.org/10.18823/asiatefl.2017.14.1.8.114>
- \*Naeini, J., & Duvall, E. (2012). Dynamic assessment and the impact on English language learners' reading comprehension performance. *Language Testing in Asia*, 2(2), 22–41. <https://doi.org/10.1186/2229-0443-2-2-22>
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3) (pp. 269–284). Lawrence Erlbaum Associates Publishers.
- Nagy, W. E., Herman, P.A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233–253. <https://doi.org/10.2307/747758>
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, P. (2019). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp.15–29). Routledge. <https://doi.org/10.4324/9780429291586>
- Nation, P. (2021). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp. 15–29). Routledge. <https://doi.org/10.4324/9781003106609-33>
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.

- Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, 10(1), 138–154. [https://doi.org/10.1044/1058-0360\(2001\)014](https://doi.org/10.1044/1058-0360(2001)014)
- Pishghadam, R., Barabadi, E., & Kamrood, A. M. (2011). The differing effect of computerized dynamic assessment of L2 reading comprehension on high and low achievers. *Journal of Language Teaching and Research*, 2(6), 1353–1358. <https://doi.org/10.4304/jltr.2.6.1353-1358>
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development* (Vol. 9). Springer Science & Business Media. <https://doi.org/10.1007/978-0-387-75775-9>
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233–265. <https://doi.org/10.1191/1362168805lr1660a>
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*, 17(3), 323–342. <https://doi.org/10.1177/1362168813482935>
- Poehner, M. E., & Wang, Z. (2021). Dynamic Assessment and second language development. *Language Teaching*, 54(4), 472–490. <https://doi.org/10.1017/S0261444820000555>
- Poehner, M. E., & Yu, L. (2022). Dynamic assessment of L2 writing: Exploring the potential of rubrics as mediation in diagnosing learner emerging abilities. *TESOL Journal*, 56(4), 1191–1217. <https://doi.org/10.1002/tesq.3098>
- \*Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337–357. <https://doi.org/10.1177/0265532214560390>
- ProQuest. (n.d.a). *ERIC (Education Resources Information Center)*. ProQuest. <https://proquest.libguides.com/eric/home>
- ProQuest. (n.d.b). *Linguistics and Language Behavior Abstracts (LLBA)*. ProQuest. <https://proquest.libguides.com/llba>
- Purpura, J. E. (2004). *Assessing grammar* (Reprint ed., Vol. 8). Cambridge University Press.
- \*Rassaei, E. (2023). The effects of text-based and audio-based dynamic glosses on L2 vocabulary learning: A dynamic assessment approach. *The Language Learning Journal*, 51(4), 509–522. <https://doi.org/10.1080/09571736.2023.2213247>
- Ratner, N. B. (2006). Evidence-based practice: An examination of its ramifications for the practice of speech-language pathology. *Language, Speech & Hearing Services in Schools*, 37(4), 257–267. [https://doi.org/10.1044/0161-1461\(2006\)029](https://doi.org/10.1044/0161-1461(2006)029)
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146–161. <https://doi.org/10.1017/S0267190504000078>
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1–32. <https://doi.org/10.1177/026553220101800101>
- Redman, S. (2003). *English vocabulary in use: Pre-intermediate and intermediate* (4th ed.). Cambridge University Press.
- Saboori, F., Pishghadam, R., Fatemi, A. H. & Ghonsooley, B. (2015). Culture and identity: Linking Iranian identity components and cultural dimensions. *Issues in Language Teaching*, 4(1), 49–78. <https://doi.org/10.22054/ilt.2015.3463>
- \*Safdari, M., Fathi, J., & Kruk, R. (2020). Investigating the role of dynamic assessment on speaking accuracy and fluency of pre-intermediate EFL learners. *Cogent Education*, 7(1), 1690232. <https://doi.org/10.1080/2331186X.2020.1818924>
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows: Size and depth of vocabulary knowledge. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 109–120. <https://doi.org/10.1017/S0261444819000326>
- Shrestha, P., & Coffin, D. (2012). Dynamic assessment, tutor mediation, and academic writing development. *Assessing Writing*, 17(1), 55–70. <https://doi.org/10.1016/j.asw.2011.11.003>
- Sternberg, R. J., & Grigorenko, E. L. (2001). All testing is dynamic testing. *Issues in Education*, 7(2), 138–170.
- \*Teo, A. K. (2012). Effects of dynamic assessment on college EFL learners' reading skills. *The Journal of Asia TEFL*, 9(1), 57–94.
- Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development*, 44(2-3), 84–97. <https://doi.org/10.1159/000057048>
- Vygotsky, L. S. (1987). *The collected works of LS Vygotsky: The fundamentals of defectology* (Vol. 2). Springer Science & Business Media.

- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.
- Vygotsky, L. S., & Rieber, R. W. (Ed.). (1998). *The collected works of L. S. Vygotsky* (Vol. 5). *Child psychology* (M. J. Hall, Trans.). Plenum Press.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13–40. <https://doi.org/10.3138/cmlr.53.1.13>
- Wilkins, D. A. (1972). *Linguistics in language teaching* (Vol. 111). Edward Arnold.
- Zeno, S., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Touchstone Applied Science Associates.
- Zhang, J., & Lu, X. (2019). Measuring and supporting second language development using computerized dynamic assessment. *Language and Sociocultural Theory*, 6(1), 92–115. <https://doi.org/10.1558/lst.31710>