English L2 vocabulary learning with clickers: Investigating pedagogical effectiveness

Anne-Marie Sénécal, Concordia University
Walcir Cardoso, Concordia University

Abstract

A growing body of literature on the pedagogical effectiveness of clickers in a second language (L2) context has revealed that clickers can promote learning. However, the extent to which clickers play a role in L2 acquisition compared to other pedagogical approaches lacks consensus; in addition, most research has focused on adult learners and has taken place in large classrooms. To address these limitations, the current research investigated the effects of clickers on L2 vocabulary acquisition in a K-12 educational setting. Two intact groups of Grade 8 students learning L2 English were assigned to a treatment: while the Clicker Group (n = 31) received instruction via clickers, the Non-Clicker Group (n = 30) was treated via hand-raising without the target technology. The pedagogical effectiveness of clickers on participants’ acquisition of the target vocabulary was measured via pretests, posttests and delayed posttests. Overall, the results indicate that vocabulary acquisition was comparable in both groups. The discussion of the findings explores the role of individual differences among users (i.e., some participants improved significantly more than others) and highlights the implications of the study for L2 teaching/learning.

Keywords: Learner Response Systems, Clickers, L2 Pedagogy, Pedagogical Effectiveness

Language(s) Learned in This Study: English


Introduction

In the last decades, the use of computer-assisted language learning (CALL) tools have been the focus of innovative pedagogical approaches to enhance students’ learning experience (Chapelle & Jamieson, 2008). However, much remains unknown about the pedagogical effectiveness of many of these CALL tools, such as learner response systems or clickers—a promising technology that has been claimed to facilitate learning (e.g., Kocak, 2022; Lee & Oh, 2014).

The beginning of the twenty-first century marked a rapid expansion of the use of clickers in educational contexts, predominantly in post-secondary institutions but also in primary and secondary education (Cardoso, 2011). In 2010, more than 8% of K-12 classrooms adopted clicker technology in the USA (Moss & Crowley, 2011). Considering the evolution of this technology and the design of a variety of types of systems, such as Bring Your Own Device web-based systems (e.g., Kahoot!, Socratic) and device-based systems (e.g., Plickers, clickers), it can be estimated that millions of students use a form of clicker technology in their classrooms.

Clickers have been implemented in pedagogical contexts to promote a variety of language features, such as vocabulary acquisition (e.g., Wichadee & Pattanapichet, 2018). There is an assumption that words can be learned implicitly through multiple encounters with words in the input and contextual guessing (e.g., Horst, 2013; Ma & Kelly, 2006; Sternberg, 1987), explicitly via instruction (Nation, 2013), or through a combination of implicit and explicit instruction (Wesche & Paribakht, 2000). Accordingly, clicker-based
instruction seems to have pedagogical potential for vocabulary acquisition, since it combines implicit and explicit features of acquisition by promoting opportunities for meaningful encounters with a word (i.e., implicit) as well as direct attention to it (i.e., explicit) (Elgort, 2018; Reynolds & Taylor, 2020).

To determine the pedagogical potential of a technological tool, Cardoso (2022) discusses a chronological framework in four stages for conducting CALL research. Step 1 involves the Development of the tool, which was completed years ago. Step 2 is the Exploration of the tool’s affordances. Research has extensively explored the affordances associated with the use of clickers, which include enhanced student participation, motivation, engagement, and interactivity (Cardoso, 2011; Kocak, 2022).

To determine if these affordances are transposed into the students’ attitudes towards the technology, the goal of Step 3 of the chronological framework is to Assess Suitability. There is a general agreement in the literature that learners have positive attitudes towards the use of clickers as an instructional approach in both general education (e.g., Oigara & Keengwe, 2013) as well as in second language (L2) contexts (e.g., Sénécal et al., 2022; Song et al., 2017).

Ultimately, once suitability has been established, it is essential to carry out Step 4 of the framework by investigating the Pedagogical Effectiveness of the tool, which corresponds to the actual learning that occurs via its implementation. Thus far, the literature suggests that the pedagogical use of clickers can benefit learning (e.g., Lee & Oh, 2014; Reynolds & Taylor, 2020). However, more research is needed to provide additional empirical evidence, particularly in K-12 contexts. Accordingly, the goal of this paper is to further explore the pedagogical effectiveness of clickers by examining the effects of this technology for L2 English vocabulary learning in a K-12 educational setting.

Background

Vocabulary

Vocabulary Knowledge

The operationalization of what it means to know a word remains unclear (Milton, 2009) due to the complex nature of the construct of vocabulary knowledge (González-Fernández & Schmitt, 2020). To simplify this concept, Ma (2009) suggests that vocabulary knowledge relates to “knowing the meaning of the word and how to use it appropriately” (p. 27). Among the many frameworks developed to operationalize vocabulary knowledge, one of the most influential is Nation’s (2013), for whom knowing a word comprises three knowledge dimensions: form (e.g., spoken and written form), meaning (e.g., definition and associations with other words), and use (e.g., grammatical function and constraints on use, based on register).

These three dimensions can each be further divided between productive and receptive knowledge (Treffers-Daller, 2020). Receptive knowledge, also conceptualized as passive knowledge (Milton, 2009), pertains to the ability to recognize the form, meaning, and use dimensions of a word in the input while reading or listening (Nation, 2013). On the other hand, productive knowledge, also identified as active knowledge, refers to the ability to convey meaning using the word via speaking and writing (Ko & Goranson, 2014). Barclay and Schmitt (2019) concede, however, that “Nation’s conceptualization should be seen as aspirational rather than prescriptive”, and that not all components of the framework are mandatory for appropriate use of vocabulary (p. 803).

Vocabulary Acquisition and Assessment

Based on this multi-faceted operationalization of vocabulary knowledge, researchers have explored vocabulary acquisition, which is defined as the increase of that knowledge over a period of time (Ma, 2009). Vocabulary acquisition may occur intentionally, with vocabulary learning as the pedagogical goal (explicit learning), or incidentally, without a conscious intention to learn (implicit learning) (Barclay & Schmitt, 2019).

A key component of implicit learning is that for learning to take place, learners need to have repeated
encounters with the target word in a variety of contexts in the input (Horst, 2013; Ma & Kelly, 2006; Nation, 2013). Using the context in which the word appears, learners can employ contextual guessing and as a result infer the meaning of unknown words (Sternberg, 1987). The exact number of encounters with a word necessary to acquire it is not definitive in the literature—Nation (1990) argues that it is between 5 and 16 encounters, while Webb (2007) posits that 10 exposures are reasonable. Frequency of exposure has been observed to be associated with vocabulary acquisition not only through implicit learning but also through explicit learning (Barclay & Schmitt, 2019). In the case of explicit learning, Peters (2014) advocates for greater exposure frequency rather than establishing a fixed number of necessary exposures.

Barclay and Schmitt (2019) argue that different components of vocabulary knowledge (see Nation’s 2013 framework above) require different pedagogical attention (implicit versus explicit). However, other researchers argue that incidental learning is not sufficient to acquire vocabulary and that explicit learning targeting specific vocabulary or the development of vocabulary learning strategies is necessary to increase the efficacy of the learning process (e.g., Horst, 2013; Ma & Kelly, 2006). Ultimately, there has been a shift in the literature to adopt a mixed approach that encompasses features of both implicit learning (e.g., via extensive reading) and explicit learning (e.g., via vocabulary-building activities) (Ma, 2009; Wesche & Paribakht, 2000).

A variety of tests have been developed to measure vocabulary learning considering that a test alone is not sufficient to measure all aspects of vocabulary knowledge (Milton, 2009), and there will be some level of overlap between the measures of vocabulary knowledge (e.g., receptive and productive knowledge) (Ma, 2009). Nation (2013) recommends question types that target different dimensions of vocabulary knowledge (i.e., form, meaning, and use), such as those that ask learners to translate the word into their first language (L1; i.e., meaning), or to create a sentence using it (i.e., use).

There has been increasing attention toward the use of CALL tools to promote vocabulary learning, particularly to support the development of implicit and explicit knowledge and related learning strategies (Ma, 2009; Taj et al., 2017). In fact, in a review of technology-mediated second language vocabulary development, Elgort (2018) states that technology has been playing and will continue to play a crucial role in vocabulary development considering the rapid sophistication and expansion of the capabilities afforded by CALL tools. The author highlights different types of technology that can promote vocabulary acquisition, such as the development of language corpora, text-to-speech technology, and the use of mobile phones (see Lin & Lin, 2019, for a systematic review and meta-analysis of mobile-assisted vocabulary learning). One such technology that has the potential to enhance vocabulary acquisition is clickers (Reynolds & Taylor, 2020).

**Clickers: Instructional Approach**

Clickers have been used as a classroom instructional approach in several fields to allow an entire group to respond to questions, usually multiple-choice questions (MCQs). This technology allows for many students to simultaneously answer questions, after which answers are wirelessly collected, transmitted to a computer, and projected on a screen. The standard procedure to use clickers pedagogically was suggested by Mazur (1997) and involves the steps outlined below and illustrated in Figure 1.

**Procedure**

**Step 1: Creating Questions.** Before the lesson, the instructor creates MCQs in the clicker software (e.g., *Turning Point by Turning Technologies*) and embeds the questions in presentation software (e.g., *Microsoft PowerPoint*). A variety of questions targeting different aspects of Bloom’s taxonomy may be designed (e.g., recognizing or recalling knowledge from memory; for a revised version of this taxonomy, see Anderson & Krathwohl, 2001).

**Step 2: Asking Questions.** At a specific time during the lesson, the instructor asks a clicker question by displaying it on the projector (Figure 1A). The instructor allocates time for students to read the question and think about their answer. Then, the instructor opens the polling, during which students answer the
question. To answer, the students press the button that corresponds to their answer on the clicker (Figure 1B). Student input is instantaneously wirelessly collected and transmitted to a receiver (Figure 1C) connected to the instructor’s computer for processing (e.g., to calculate answer distributions and other statistical information).

**Step 3: Providing Feedback.** Once all answers are recorded, the instructor closes the polling. On the instructor’s computer (and optionally on the projected screen), the distribution of student answers appears in a visual format (Figure 1D). Lasry et al. (2008) proposes a Peer Instruction Implementation Process to help instructors proceed depending on the answer distribution. If the graphic shows that most students chose the correct answer (>70%), the instructor should display the answer distribution graphic, show the correct answer, and move on. If the graphic shows that students are divided in their answers (30-70%), the instructor should prompt a convince-your-neighbour discussion before showing the answer distribution graphic. During this discussion, students work in pairs or in small groups and provide rationales for their answer to try to convince each other of what the correct answer is. A new polling session can follow the discussion, which allows the instructor to reassess the responses. Finally, if only very few students have the correct answer (<30%), the instructor should provide further explanation of the topic. Ultimately, the answer distribution graphic informs both the students and the instructor of the level of understanding of the target question and may influence how the instructor proceeds, thus enabling contingent teaching (Cardoso, 2013; Lasry et al., 2008). Accordingly, the proposed teaching intervention adopts aspects of Test, Teach, Test, where Test is when learners first answer clicker questions. Teach can take place after the students’ level of understanding is revealed by the answer distribution graphic. Based on the students’ needs, teachers can adapt their teaching by providing additional instruction on the topic or prompting peer instruction through a convince-your-neighbour discussion. Finally, Test occurs when the clicker questions are prompted a second time after explicit teaching or peer instruction. In this context, the Test, Teach, Test approach can occur iteratively, with multiple cycles of the last two steps.

**Figure 1**

*Procedure to Use Clickers*

<table>
<thead>
<tr>
<th>A</th>
<th>Which of these is a synonym of the word <strong>dainty</strong>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Repulsive</td>
<td>b) Delicate</td>
</tr>
<tr>
<td>c) Soft</td>
<td>d) Large</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image of clicker" /></td>
<td><img src="image2.png" alt="Image of clicker" /></td>
<td><img src="image3.png" alt="Image of clicker" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image of clicker" /></td>
<td></td>
<td>Which of these is a synonym of the word <strong>dainty</strong>?</td>
</tr>
<tr>
<td>a) Repulsive</td>
<td>b) Delicate</td>
<td></td>
</tr>
<tr>
<td>c) Soft</td>
<td>d) Large</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Repulsive</th>
<th>Delicate</th>
<th>Soft</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>30%</td>
<td>25%</td>
<td>15%</td>
</tr>
</tbody>
</table>

**Pedagogical Affordances**

The rapid expansion of the use of clickers in educational contexts may be associated with the affordances of the technology, which consolidate current educational theories. One affordance is that the technology fosters opportunities for peer instruction, a student-centered approach that is assumed to promote learning (Mazur, 1997), particularly via convince-your-neighbour discussions (Lasry et al., 2008). Peer instruction also creates a collaborative learning environment, which increases opportunities for interactions between learners (Hake, 1998) as well as between learners and technology (Chapelle, 2003).

Another important affordance of clickers is that the answer distribution graphic compiled from the polling results provides immediate feedback (Hattie & Timperley, 2007). For students, seeing the distribution of
answers allows them to self-assess their understanding and to compare it to that of the other students (Draper et al., 2002; Kocak, 2022). For instructors, the immediate feedback offered (e.g., in formative assessments) can uncover prevailing misconceptions, which in turn can inform teaching by allowing instructors to instantaneously customize a lesson (Caldwell, 2007).

An additional affordance is that the environment fostered by clickers promotes student engagement and participation (Oigara & Keengwe, 2013). Since all students must press a button to answer each question, as opposed to a small group of students raising their hand to answer the question, learning attention is improved and participation increased (Hunsu et al., 2016). Student participation is further incited by the fact that when answering questions with clickers, the process is anonymous, which helps promote participation by reducing the fear of peer criticism or losing face in front of colleagues (Caldwell, 2007; Ryoo, 2015).

Overall, these affordances highlight that clickers can be used as active-learning tools (see Kozanitis & Nenciovici, 2022 for a meta-analysis on the positive pedagogical effects of active learning in comparison with traditional lecturing). These pedagogical benefits have motivated researchers to empirically investigate if the affordances fostered by the technology can translate into pedagogical effectiveness.

**Pedagogical Effectiveness in General Education**

Clicker technology has increasingly been integrated into classrooms in a variety of fields and disciplines (Cardoso, 2011; 2013). However, a large portion of the literature on this tool has investigated its use in the fields of science, technology, engineering, and math (STEM fields)—particularly in biology (Pan et al., 2019) and math (Stewart & Stewart, 2013). Research in other educational contexts has also been carried out, such as in psychology (Iwamoto et al., 2017) and educational technologies (Ranieri et al., 2018).

Researchers have also explored specific aspects of the use of clickers in general education. For instance, Wang (2015) investigated the possibility of a wear-off effect, also known as a novelty effect, between two comparison groups (i.e., using the technology during one lecture versus an entire semester) in a study of 252 university-level students from a variety of fields. The author found that, although there was a minimal novelty effect after a longer integration of Kahoot! over a five-month semester, the impact was not substantial and did not affect the pedagogical effectiveness of the tool.

Overall, there is a general agreement that the pedagogical use of clickers promotes learning (e.g., Iwamoto et al., 2017; Pan et al., 2019). However, the literature does not reach the same level of agreement as to the extent that clickers generate more learning when compared to other non-clicker pedagogical approaches (e.g., hand raising or other traditional voting methods). For example, Iwamoto et al. (2017), in a study of university students enrolled in a psychology course, found that using Kahoot! (a clicker-based learning platform) in the last 10 minutes of the class yielded significantly more learning gains compared to students in a control group not using the technology. Contrarily, other studies do not show significantly higher learning gains in clicker-enhanced learning. Consider Ranieri et al. (2018), for instance, who explored the pedagogical effectiveness of Kahoot! using a within-subjects design among university-level students in large classes in an educational technology course. The authors found comparable learning gains between the experimental lessons in which Kahoot! was used and the control lessons not using the technology (Ranieri et al., 2018).

**Pedagogical Effectiveness in L2 Acquisition**

More recently, a growing body of literature has explored the pedagogical effectiveness of clickers focusing on various components of L2 learning. Consistent with the results from general education settings, there is an agreement in past studies that clickers foster learning gains in an L2 context, but there is a disagreement when comparing pedagogical approaches (e.g., Baran-Lucarz et al., 2015; Yu & Yu, 2017). Some of these studies reveal that clicker-enhanced learning leads to significantly higher learning gains. For example, Lee and Oh (2014) found that, in a university-level reading comprehension course, the experimental group using clickers generated significantly higher results than the control group not using the technology. Similar
results were obtained by Agbatogun (2014), who investigated the effects of clicker technology on grade six students’ communicative competence. The author found that the experimental group that utilized clickers had significantly higher communication gains compared to both the comparison group that employed a communicative pedagogical approach and the control group that followed lectures.

In a flipped classroom context, Hung (2017) explored the use of Kahoot! to promote speaking skills between two comparison groups, one using Kahoot! as a teacher-centered questioning approach, and another group using the technology with convince-your-neighbour discussions. The results found that the experimental group using Kahoot! had significantly higher developments in speaking skills.

However, Baran-Lucarz et al. (2015) uncovered mixed findings in a study on the effects of clickers in a university-level English L2 phonetics course between an experimental group using clickers and a control group not using the technology but the same curriculum. They found significantly more learning gains in the experimental group for some measures (e.g., stressing words in reading aloud tasks), but not in all measures considered (e.g., predicting word stress patterns), thus reinforcing the assumption that the pedagogical effectiveness of clickers remains unclear in the literature.

**Pedagogical Effectiveness in L2 Vocabulary Acquisition**

The same pattern of results can also be observed in the literature specifically focusing on L2 vocabulary acquisition. Previous studies agree that the pedagogical use of clickers can promote learning (e.g., Guaqueta & Castro-Garcés, 2018) but disagree on the extent to which clicker-enhanced learning leads to higher learning gains when compared with other pedagogical approaches. For example, Reynolds and Taylor (2020) explored vocabulary acquisition in a university-level English L2 course and observed comparable acquisition between the experimental group using Kahoot! to reinforce the coursebook’s target lexis and the control group solely using the coursebook. Contrarily, Wichadee and Pattanapichet (2018) found significantly higher learning gains in the experimental group using Kahoot!-based vocabulary quizzes than in the control group using a paper version of the quizzes.

**Current Study and Research Questions**

Overall, previous studies show that the pedagogical use of clickers can foster learning gains (e.g., Lee & Oh, 2014; Wichadee & Pattanapichet, 2018). However, the extent to which clickers play a role in learning gains compared to other pedagogical approaches lacks consensus. Further, most research has focused on adult learners and has taken place in large classrooms. The goal of this study is to address these gaps by providing additional empirical evidence regarding the Pedagogical Effectiveness of clickers (Step 4 of Cardoso’s 2022 chronological framework for assessing a tool’s pedagogical potential) in an L2 learning context in a K-12 educational setting. The following research question guided the study:

- To what extent does the pedagogical use of clickers contribute to the acquisition of L2 English vocabulary in a K-12 educational context?

Despite the inconclusive results observed in the literature in terms of the tool’s pedagogical effectiveness, we hypothesize that students learning with clickers will outperform those not using it due to the benefits afforded by the technology, as discussed earlier (e.g., it promotes peer instruction, provides quality feedback, and increases engagement).

**Methods**

**Participants**

Sixty-one student participants were recruited from a private francophone high school in Montréal, Québec (Canada). Participants were in grade eight (13-14 years old). French was the participants’ dominant language. While the vast majority of students used French as their L1, four considered themselves plurilingual, speaking French and another language (i.e., English, Mandarin Chinese, and Russian). The participants were enrolled in the ministry-mandated Core ESL program, which indicates they were high-
A convenience sampling method was employed to recruit the participants who constituted two intact groups taught by the same English teacher and had no prior experience using clickers. The groups were randomly assigned to an experimental group: (a) the Clicker Group (CG; \( n = 31 \)), in which students received vocabulary instruction via clicker technology; and (b) the Non-Clicker Group (NCG; \( n = 30 \)), in which vocabulary instruction was carried out without the clickers (e.g., via hand-raising and other traditional methods). Participation in the study abided by standard ethical procedures and did not count for course grades.

**Materials: Vocabulary Instruction Treatment**

Over eight weeks, students took part in vocabulary-building activities that targeted words extracted from a required novel for the year, which allowed the integration of the treatment into the regular course curriculum. This approach for the treatment was chosen as Barclay and Schmitt (2019) acknowledge that recent research shows reading contributes to the acquisition of vocabulary. The novel, *James and the Giant Peach* by Dahl (1996), was selected as it was deemed age- and level-appropriate for grade eight students by the teacher. Thirty words were extracted from this novel to assess the students’ vocabulary knowledge before and after the treatment. The words included 10 nouns, 10 adjectives, and 10 verbs. Although some words carry two distinct content word classes (e.g., *gaze* can be a noun and a verb), the treatment focused on the class in which the word appeared in the book (input).

The lexis was selected to promote vocabulary acquisition because of the instruction method rather than due to encounters with the words outside of the treatment. Accordingly, to limit students’ exposure to the items outside of the study, high-frequency vocabulary, which consists of the most frequent 2,000-word families, were excluded (Nation, 2013). See Appendix A for the list of vocabulary items.

The book was divided into five reading sections (roughly 30 pages per section), which were each followed by an in-class vocabulary instruction session that lasted approximately 30 minutes. Each weekly session targeted six words that corresponded to the previous reading section. For example, the first reading section included chapters 1-8. In this reading, students encountered the word *dainty* at least once (chapter 2). Therefore, the word *dainty* was targeted in the first vocabulary instruction session. The vocabulary instruction sessions were carried out by the teacher, during which students answered MCQs to reinforce their acquisition of the words. It is important to note that although the teacher carried out other learning activities about the target novel with the students, the scope and structure of these activities were not disclosed to the researchers, as they were deemed beyond the scope of the current project.

In the CG, the MCQs were created in a clicker software and then embedded in PowerPoint presentations. To answer the questions, students anonymously responded using clickers. Although the popularity of Bring Your Own Device tools has been steadily increasing (Reynolds & Taylor, 2020), we opted for a device-based technology to accommodate our teaching context: high school students. Since we could not assume that all students own or have access to personal electronic devices, using device-based clickers was the optimal solution. Also, in some institutions, students are not allowed to carry their personal electronic devices in the classroom. In the NCG, the MCQs were included in the PowerPoint presentation, and the participants answered via hand-raising (as is customary in their classes). Therefore, the treatment was identical in both the CG and the NCG groups; the only difference was that students in the CG responded to questions using their clickers.

These MCQs, designed by the research team prior to the vocabulary instruction sessions, were based on the context in which the words appeared in the novel. For each word, the first slide included the passage from the novel in which the word appeared. In the passage, the word was bolded, in italics, and in colour to increase saliency. Then, two MCQs directly addressed the understanding of the target word (e.g., by selecting the most appropriate definition, translation, or synonym for the word). In the question, the target word appeared in italics to increase saliency. Figure 2 illustrates the PowerPoint sequence for the word *dainty*. 
During the vocabulary-instruction sessions, convince-your-neighbour activities (see also Procedure for details) took place to promote peer instruction (i.e., explicit learning). In the CG, the discussion was prompted based on Lasry et al.’s (2008) procedure (e.g., a spontaneous discussion was prompted if 30% or more students select an incorrect response). In the NCG, similar activities were also implemented, but without the clicker affordances (e.g., anonymous voting). During these discussions, students were encouraged to employ learning strategies such as using context cues (e.g., to infer the meaning of the target word). Learning strategies are part of the Québec ESL program and it is expected that by grade eight (i.e., the level of our participants), learners should be able to infer meaning as a vocabulary-learning strategy.

Overall, the treatment was aligned with research that supports a mixed approach for vocabulary instruction, combining features of both implicit and explicit learning (Ma, 2009; Wesche & Paribakht, 2000). For instance, learners incidentally encountered each target word in the input (i.e., implicit learning) when reading the book and during the vocabulary-building activities (i.e., in the MCQs themselves and during the convince-your-neighbour discussions about the lexis). During the vocabulary-building activities, students answered two clicker-based MCQs per item, which provided additional practice to foster vocabulary acquisition (i.e., explicit learning). Hence, through the treatment, the learners had repeated encounters with each word both implicitly and explicitly by reading it, seeing it, and discussing it to reinforce the acquisition of the target items, as recommended by Horst (2013) and Nation (2013).

**Figure 2**

_Vocabulary Instruction Session: PowerPoint Presentation Sequence for a Word (Dainty)_

1. **Slide 1:** Passage from the novel in which the word appears

   _And if I take off both my socks_  
   _You'll see my _**dainty**_ toes._  

2. **Slides 2-3:** Two MCQs to foster the acquisition of the target word

   - Which of these is a synonym of the word _dainty_?  
     - a) Repulsive  
     - b) Delicate  
     - c) Soft  
     - d) Large

   - Which sentence uses the word _dainty_ most appropriately?  
     - a) The ground was dainty after the heavy rain.  
     - b) The smell of rotten eggs was dainty.  
     - c) They served dainty pastries on a silver platter.  
     - d) My brother has a dainty appetite.

**Instruments: Vocabulary Assessment**

The students were allocated one hour to complete each assessment or test, but the majority finished within 30 minutes. The vocabulary assessment was administered on three occasions: (a) pretest, to measure the participants’ initial knowledge of the target words before the treatment; (b) immediate posttest, to measure the participants’ knowledge of the target words (i.e., the number of words learned) immediately after the treatment; and (c) delayed posttest, to determine the lasting effect of the treatment four weeks later. The
assessment was identical on all three occasions, and it included the 30 vocabulary words from the treatment as well as six distractors. The same assessment was administered in the CG and in the NCG. The types of questions were similar to the MCQ format (i.e., selected-response item) used during the instruction sessions, but they required students to instead produce their own answers (i.e., constructed-response item). For this task, the participants were asked to demonstrate their knowledge of the words through various means: (a) write a definition, (b) create a meaningful sentence using the word (in their L1 or target L2), (c) draw a picture representing the word, or (d) translate the word to French (the participants’ language of education). The format of the vocabulary assessment was adapted from Cardoso’s (2011) study. As discussed earlier, Nation (2013) recommends an assessment that targets different dimensions of vocabulary knowledge (i.e., form, meaning, and use). In this assessment, meaning (receptive skills) can be demonstrated via the definition, the picture, and the translation, whereas use can be shown through the creation of a sentence.

**Procedure**

A pretest-posttest-delayed posttest research design was adopted to measure the pedagogical effectiveness of the vocabulary instruction treatment on the participants’ acquisition of the 30 target vocabulary words with clickers (CG) and without clickers (NCG). The following outlines the key components of the procedure.

During **Week 1**, the researcher visited the school twice. On the first visit, signed consent forms (by the students and their parents) were gathered, and the participants completed a biodata survey as well as the vocabulary assessment pretest. On the second visit, the researchers carried out a training session to familiarize the teacher and the students with the use of clickers. During that session, the teacher was given instructions about the pedagogical use of the technology and Lasry et al.’s (2008) procedure for convince-your-neighbour discussions.

Between **Weeks 2-8**, the vocabulary instruction treatment took place in the two experimental groups. During the treatment, the participants read the novel *James and the Giant Peach* as homework assignments and took part in the five in-class vocabulary instruction sessions to learn the 30 words.

During **Week 8**, the research team visited the school again to administer the vocabulary assessment posttest immediately following the final vocabulary instruction session. A student perception survey was also administered (not reported here).

During **Week 12**, four weeks after the end of the treatment, the researchers visited the school for the last time to administer the vocabulary assessment delayed posttest. No vocabulary-building activities for the target lexis were assigned to the students between the immediate and delayed posttests. Interviews were also carried out with the students and teacher. Due to the scope of this study, user perceptions will not be discussed. However, these results revealed that while student-participants had positive attitudes towards clickers, the teacher had neutral to negative perceptions towards the tool (see Sénécal et al., 2022). **Figure 3** provides a visual representation of the procedure.
Data Collection and Analysis

The quantitative data collected to measure the students’ acquisition of the target vocabulary items included the responses for each item on the vocabulary assessments. A score of 1 was awarded if the response showed understanding of the word; otherwise, a score of 0 was given. Two researchers individually scored the three tests (pretest, posttest, delayed posttest). In case of scoring disparities \( n = 25/216 \text{ answers} \approx 11.6\%; \) for the three testing times in both groups), the researchers reached a verbal agreement. A final score out of 30 was obtained for each test. The numeric test scores were then entered into the statistical program SPSS (25.0).

Preliminary data analyses were conducted to measure the reliability of the Vocabulary Assessment tests at three times via Cronbach’s alpha. The results showed that the pretest had a reliability coefficient of .633, that the immediate posttest had a reliability coefficient of .694, and that the delayed posttest had a reliability coefficient of .679. Thus, the instruments had an adequate level of internal consistency (Nunnally & Bernstein, 1994).

The pretest-posttest-delayed posttest final scores were analyzed using descriptive statistics (i.e., mean, median, and standard deviation). Assumptions of normality of distribution and homogeneity of variance were assessed according to Field (2018) before parametric tests were conducted. An independent-samples \( t \) test was then used to compare the participants’ initial level of vocabulary between the experimental groups at the pretest (i.e., the number of words the participants knew before starting the treatment). Boxplot assessment identified one outlier in the NCG at the pretest (i.e., showing higher vocabulary knowledge),
but there was no reason to remove the participant’s data from the results. Pretest scores on the Vocabulary Assessment for each group were not normally distributed, as assessed by Shapiro-Wilk test (p > .05). Considering the robustness of t-tests and since sample sizes are nearly equal between the groups (CG: n = 31; NCG: n = 30), violations of normality would not cause problems. Finally, there was homogeneity of variances, as assessed by Levene’s test for equality of variances (p = .279).

The data were analyzed using a mixed between-within ANOVA to determine differences between test scores at times of testing and between experimental groups, as well as to determine whether there was an interaction between time and group. Boxplot assessment did not identify any extreme outliers in the immediate posttest and delayed posttest. Assessed by the Shapiro-Wilk test, the data deviated from a normal distribution (all ps < .05), except for the CG at the posttest (p = .207). Considering the robustness of ANOVAs, violations of normality would not cause problems. The assessment of Box’s Test of equality of covariance matrices was violated (p = .007). Although Box’s test was significant, Hotelling’s Trace is robust in the two-group situation; therefore, we can trust the mixed ANOVA results (see Field, 2018, p. 753). The assumption of sphericity was violated as assessed by Mauchly’s test of sphericity, χ²(2) = 8.03, p = .018. Therefore, a Greenhouse-Geisser correction was applied to the degrees of freedom for the main effect of time and those of the interaction effect (ε = 0.886). As assessed by Levene’s test for equality of variances, homogeneity of variances was met at the pretest, F(1, 59) = 1.19, p = .279; however, it was violated at the immediate posttest, F(1, 59) = 10.61, p = .002, and at the delayed posttest, F(1, 59) = 6.54, p = .013. Given that the groups can be considered equal in size, Hartley’s Fmax test using the variance ratio for these two tests showed that this is not a concern. In the case where the mixed ANOVA results were significant, Bonferroni post hoc t tests were applied to allow for a more conservative estimation of the statistical significance (p at .05/3 = .0167).

Results

The descriptive statistics illustrated in Table 1 show an overview of student performance in both groups at the three testing times. Before analyzing vocabulary acquisition, an analysis of pretest results was run to compare the participants’ initial vocabulary knowledge of the target 30 words before the vocabulary instruction treatment. An independent-samples t test revealed no difference between the CG (M = 1.29, SD = 1.77) and the NCG (M = 0.93, SD = 1.39), t(59) = 0.87, p = .39. Thus, both groups had comparable vocabulary knowledge of the target words at the pretest.

Table 1

<table>
<thead>
<tr>
<th>Test</th>
<th>Clicker Group (CG)</th>
<th>Non-Clicker Group (NCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 31</td>
<td>n = 30</td>
</tr>
<tr>
<td></td>
<td>M/30</td>
<td>SD</td>
</tr>
<tr>
<td>Pretest</td>
<td>1.29</td>
<td>1.77</td>
</tr>
<tr>
<td>Immediate Posttest</td>
<td>6.61</td>
<td>3.81</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>4.29</td>
<td>3.09</td>
</tr>
</tbody>
</table>

To measure the level of significance of vocabulary learning, a statistical analysis of the data was conducted via a mixed between-within ANOVA. The results indicated a significant difference in vocabulary knowledge for the main effect of time, $F(1.77, 104.50) = 159.53, p < .001, \eta_p^2 = .730$, but there was not a significant difference for the main effect of group, $F(1, 59) = 2.78, p = .101, \eta_p^2 = .045$, and there was no significant Group × Time interaction, $F(1.77, 104.50) = 3.03, p = .052, \eta_p^2 = .049$.

Post hoc analysis of the significant effect of the within-subjects variable time with a Bonferroni adjustment...
revealed that there was significant improvement in learning between the pretest and immediate posttest, $M_{\text{difference}} = 4.68$, 95% CI [3.93, 5.43], $p < .001$, $d_z = 1.87$, with a large effect size (Field, 2018, p. 115), but a significant decrease was observed between the immediate posttest and delayed posttest, $M_{\text{difference}} = -1.95$, 95% CI [-2.51, -1.38], $p < .001$, $d_z = 0.67$, with a medium effect size (Field, 2018, p. 115). However, the improvement between the pretest and delayed posttest remained statistically significant, $M_{\text{difference}} = 2.73$, 95% CI [2.12, 3.35], $p < .001$, $d_z = 1.25$, with a large effect size (Field, 2018, p. 115).

Altogether, the analyses indicated that both groups had comparable improvement over time. It is also notable that higher standard deviations were observed in the CG at the immediate posttest (CG: $SD = 3.81$; NCG: $SD = 2.31$) as well as at the delayed posttest (CG: $SD = 3.09$; NCG: $SD = 2.08$). See Figure 4 for an overview of the group comparisons in vocabulary learning.

Figure 4

Performance on Vocabulary Assessment Tests

Discussion

To examine the pedagogical effectiveness of clickers for L2 vocabulary acquisition, the study set out to answer the following research question: to what extent does the pedagogical use of clickers contribute to the acquisition of L2 English vocabulary in a K-12 educational context?

To answer the question and thereby measure the participants’ vocabulary acquisition, we analyzed the vocabulary assessments administered in the CG and the NCG before (pretest) and after (immediate and delayed posttests), which focused on the 30 vocabulary items targeted during the treatment. Overall, the results indicate that the proposed treatment contributed to the learning of the target English vocabulary in both the CG (via clickers) and the NCG (without clickers), but that the vocabulary acquisition in the group with clicker-enhanced instruction (i.e., CG) was not statistically higher than those observed in the group not using the technology (i.e., NCG).
These findings suggest that, while the pedagogical use of clickers may promote the acquisition of L2 vocabulary, different pedagogical approaches that target L2 vocabulary may have an equivalent effect. Considering that the literature is inconsistent regarding the extent to which the use of clickers leads to higher vocabulary acquisition when compared with traditional methods of teaching (e.g., hand-raising techniques), the findings reported here lend support to the claim that clicker-enhanced instruction does not always yield more pedagogical gains (e.g., Reynolds & Taylor, 2020 for vocabulary). Accordingly, it is possible that the benefits afforded by clickers may lay elsewhere (e.g., they create a game-like, motivating environment for learning to occur; Cardoso, 2011).

The lack of a significant difference in the L2 vocabulary improvement between the CG and the NCG may be associated with the design of the treatment. As discussed in this study’s methodology, as an attempt to create a control group in which the only different variable was the use of the clicker device, characteristics inherent to the CG treatment were artificially applied in the NCG. As such, one could even claim that the NCG also received a type of clicker-assisted pedagogy, which may have inadvertently boosted learning in the NCG. For example, in the CG, small-group convince-your-neighbor discussions were strategically prompted in situations where the answer distribution graphic (i.e., a fundamental feature of clicker-enhanced instruction) showed a disparity in student answers. These discussions were artificially implemented in the NCG: without an answer distribution graphic indicating whether the students had indeed learned the word being taught, a post-MCQ session discussion was arbitrarily prompted (as dictated by this study) in that group. In practice, teachers in settings in which clicker technology is not available would rarely opt for these types of small-group discussions without the cues afforded by the answer distribution graphic.

As a result of the design of the treatment, similar potential for vocabulary acquisition was afforded in both groups, wherein a combination of explicit and implicit approaches to teaching vocabulary was adopted (see Wesche & Paribakht, 2002 and Reynolds & Taylor, 2020 for the rationale). In both the CG and NCG, vocabulary-building activities in the form of MCQs introduced learners to target words in the context in which they appeared in the book (i.e., implicit learning) and prompted questions that targeted different aspects of vocabulary knowledge (i.e., explicit learning). For instance, to develop the participants’ lexical repertoire, the proposed vocabulary-building activities targeted Nation’s (2013) three dimensions of vocabulary knowledge (i.e., form, meaning, and use) as well as Ma’s (2009) receptive and productive components of what it constitutes to know a word. Further, the MCQs provided students with multiple opportunities to encounter the word in context (Nation, 2013; Webb, 2007). In both treatment groups, repeated exposure to the target lexis was promoted via the small-group discussions, which were prompted strategically in the CG (i.e., based on the answer distribution graphic) and artificially in the NCG. It could be argued that the similarities in the pedagogical approach adopted for both CG and NCG could partially account for the comparable vocabulary acquisition in the two groups.

Relatively similar assumptions have also been reported in the clicker literature. In a study on L2 learner perceptions of clickers, for instance, Cardoso (2011) concluded that his participants’ favorable attitudes towards the technology may be attributed to characteristics inherent to the pedagogical approach that the technology affords (e.g., small-group discussions, input enhancement via the display of the target word on the screen and a graphic indication of the correct answer) rather than to the clicker itself. This analysis can also be applied to this study, mutatis mutandis, considering that the two experimental groups received a certain type of clicker-assisted pedagogy, which consisted of many of the pedagogical affordances that the technology provides.

The lack of difference in vocabulary learning between the CG and the NCG may also be attributed to the presence of individual differences in the acquisition process. In a review of the literature on the effects of clickers on student performance, Landrum (2015) suggests that factors related to individual differences may explain the inconsistent acquisition findings from clicker-enhanced instruction. For example, Roth (2012) explored the effects of aptitude as an individual difference that may affect the learning process when using clicker-enhanced instruction in a calculus course. The author found that lower-performing students (i.e.,
the students below the median score on a previous in-class evaluation) benefited more from the treatment than the higher-performing students. Unfortunately, the role of individual differences in CALL remains an under-investigated topic and has only recently received some attention in the literature (Landrum, 2015).

In this study, at the time of the immediate posttest, higher levels of standard deviation were observed among learners in the CG ($M = 6.61, SD = 3.81$) than those in the NCG ($M = 4.97, SD = 2.31$), thus creating a large disparity in vocabulary acquisition between learners. This indicates that while many participants in the CG highly benefitted from clicker-enhanced instruction, this was not true for some. For example, consider Jordan and Sacha (fictitious names), two participants who exhibited extreme disparities in results (thus contributing to high standard deviations in the results obtained). While Jordan considerably benefitted from clicker-enhanced instruction, learning 10 vocabulary items between the pretest (0/30 score) and delayed posttest (10/30 score), Sacha did not benefit as much, as she only learned one word between the pretest (1/30 score) and delayed posttest (2/30 score). The presence of such disparities in vocabulary learning among learners in the CG (also evidenced by high standard deviations) demonstrate the presence of individual differences in the acquisition process, indicating that the proposed clicker-enhanced vocabulary instruction treatment was more beneficial for some students than others.

**Limitations**

This study has uncovered methodological limitations that should be recognized and considered for future research. Primarily, the target vocabulary selected for the treatment and tests is an aspect of the research design that challenges the implications of the results. Even though some target words had multiple word classes (e.g., gaze is both a noun and a verb), the treatment focused solely on the class in which it appeared in the book (a noun in this case). Moreover, in the study, low-frequency vocabulary (i.e., not within the most frequent 2,000-word families) from Nation’s (2013) word lists were chosen as an attempt to limit exposure to the target words beyond the scope of the project and to control for the frequent words that participants could have learned prior to the study. As the learning burden of the target lexis was increased (Webb & Nation, 2017), the choice of words may have inadvertently limited learning gains. A word’s learning burden is associated with the amount of effort required for its acquisition and depends on a variety of factors, such as the amount of exposure to and usefulness of a word (Nation, 2013). This pattern was observed in both groups. For instance, in the CG, students correctly defined 6.61 words out of 30 on the posttest compared to an average of 1.29 on the pretest. Thus, the results show that between times of testing, participants had a relatively low amount of vocabulary retention, which may be associated with the high learning burden of the low-frequency target lexis (Nation, 2013).

The sampling approach is another methodological limitation of the project. In fact, the sample of participants was limited ($n = 61$) and thus increases type II error concerns. However, Gillespie (2020) notes that empirical studies in CALL tend to be of small scale (e.g., some with fewer than 10 participants). Moreover, the participants were recruited from only one institution (one high school in Montréal, Québec). In a methodological synthesis focusing on sampling, Vitta et al. (2021) highlight the external validity limitations associated with single-site samples, regardless of the sample size. Moranski and Ziegler (2021) concede, however, that the majority of recent L2 research still occurs in single-site contexts due to a series of obstacles associated with multisite research (e.g., funding, logistics).

Finally, the fact that the target language of instruction is English constitutes another methodological limitation of the study. As Gillespie (2020) highlights, an overwhelming majority of empirical studies in CALL focus on English as the target language of instruction. Accordingly, the author calls for research to investigate under-researched target languages.

**Implications for L2 Pedagogy**

Despite these limitations, the results of this study have implications for L2 teaching and learning. Teachers should consider implementing this CALL tool in their L2 classroom since clickers not only show promising results regarding their pedagogical effectiveness, but they also promote foreign/second language enjoyment, another crucial dimension of L2 acquisition to cater to students’ affective factors (for the
rationale, see Zeng, 2021). In fact, clicker-enhanced instruction has been shown to promote engagement and motivation (Caldwell, 2007; Hunsu et al., 2016; Oigara & Keengwe, 2013), as well as to provide opportunities for peer instruction (Mazur, 1997) via small-group discussions and immediate feedback (Hattie & Timperley, 2007) through the answer distribution graphic. However, the individual differences discovered in the CG at the posttest suggest that some learners benefit more than others from this pedagogical approach. Accordingly, another essential implication for L2 pedagogy is that teachers should vary instructional approaches to capitalize on the affordances of clicker pedagogy (e.g., answer distribution graphic) to accommodate students’ needs, interests, and individual differences. Thus, clickers can be considered an essential tool in an ESL teacher’s toolkit to promote L2 vocabulary acquisition.

Conclusion

The purpose of this study was to examine the pedagogical effectiveness of clickers in a K-12 education context to foster the development of L2 vocabulary acquisition. Overall, the results indicate that the two participant groups (i.e., Clicker Group and Non-Clicker Group) had significant gains in vocabulary acquisition. However, the group that received clicker-informed pedagogy (including the use of clickers and/or the techniques associated with them) did not have learning gains statistically higher than those in the group that did not use the target technology.

Future research should expand the present study to further explore the pedagogical use of clickers. To address limitations associated with the sampling method, Moranski and Ziegler (2021) recommend that future research should consider replicating existing literature conducted with single-site sampling to provide, to a certain extent, the same effect as multisite sampling. Accordingly, we encourage future research to replicate this project and expand its pool of participants by (a) recruiting a larger number of L2 learners, and (b) conducting multisite research. Future studies should also consider exploring the pedagogical effectiveness of the technology on different L2 features other than vocabulary acquisition as was done in this project (e.g., a grammar concept). Moreover, researchers should compare different types of clicker technology (e.g., Zoom poll, Plickers) to explore the affordances inherent to the different applications. For example, Plickers is a clicker-type tool that has a unique feature: the instructor can create a MCQ spontaneously during a lesson and launch it simultaneously to gather student responses. Finally, researchers should further investigate this technology to better understand its full potential in L2 classrooms (e.g., considering its pedagogical affordances such as contingent teaching and its inherent game-like mechanics) and design an approach for clicker-enhanced instruction that could help stakeholders successfully implement the technology.

Acknowledgements

We would like to thank Turning Point Canada and the participants from Collège Sainte-Anne. Funding for this study was provided by the Social Sciences and Humanities Research Council of Canada.

References


Gillespie, J. (2020). CALL research: Where are we now? ReCALL, 32(2), 127–144. https://doi.org/10.1017/S0958344020000051


### Appendix A. Target Vocabulary Words for Instruction and Testing

<table>
<thead>
<tr>
<th>Vocabulary Instruction Sessions</th>
<th>Vocabulary Words with Parts of Speech</th>
<th>Nouns</th>
<th>Adjectives</th>
<th>Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gaze</td>
<td>Dainty</td>
<td>To beckon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seesaw</td>
<td>Stale</td>
<td>To glimpse</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Flock</td>
<td>Steep</td>
<td>To mutter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dew</td>
<td>Damp</td>
<td>To shriek</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Burden</td>
<td>Aghast</td>
<td>To shrivel</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pandemonium</td>
<td>Scruptious</td>
<td>To wail</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Blunderhead</td>
<td>Rambunctious</td>
<td>To quiver</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Candyfloss</td>
<td>Eerie</td>
<td>To gasp</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Smitherens</td>
<td>Tremendous</td>
<td>To glare</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nursery</td>
<td>Eager</td>
<td>To haul</td>
<td></td>
</tr>
<tr>
<td>Control words</td>
<td>Pail</td>
<td>Ramshackle</td>
<td>To flap</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nincompoops</td>
<td>Gloomy</td>
<td>To lurk</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Control words were not included in the vocabulary instruction sessions. They only appeared on the vocabulary tests.

### About the Authors

Anne-Marie Sénécal (MA from Concordia University) is a graduate student pursuing her PhD in Education at Université de Sherbrooke. Her research interests include the pedagogical use of clickers, educational games, and speech technologies (TTS, ASR) in L2 education. She also teaches ESL at École de musique Vincent-d’Indy, a college in Montréal (Canada). Anne-Marie Sénécal is the corresponding author.

**Email:** anne-marie.senecal@usherbrooke.ca

**ORCiD:** https://orcid.org/0000-0001-8725-0413

Walcir Cardoso is a Professor of Applied Linguistics at Concordia University. He conducts research on the L2 acquisition of phonology, morphosyntax and vocabulary, and the effects of computer technology (e.g., clickers, text-to-speech synthesizers, automatic speech recognition, intelligent personal assistants) on L2 learning.

**E-mail:** walcir.cardoso@concordia.ca

**ORCiD:** https://orcid.org/0000-0001-6376-185X