Effects of digital dictionary format on incidental acquisition of spelling knowledge and cognitive load during second language learning: Click-on vs. key-in dictionaries

Tzu-Chien Liu,*, Melissa Hui-Mei Fan, Fred Paas

Article info
Article history:
Received 3 May 2013
Received in revised form 25 July 2013
Accepted 6 August 2013

Keywords:
Evaluation of CAL systems
Media in education
Human–computer interface

Abstract
Recent research has shown that students involved in computer-based second language learning prefer to use a digital dictionary in which a word can be looked up by clicking on it with a mouse (i.e., click-on dictionary) to a digital dictionary in which a word can be looked up by typing it on a keyboard (i.e., key-in dictionary). This study investigated whether digital dictionary format also differentially affects students’ incidental acquisition of spelling knowledge and cognitive load during second language learning. A comparison between a click-on dictionary condition, a key-in dictionary condition, and a non-dictionary control condition for 45 Taiwanese students learning English as a foreign language revealed that learners who used a key-in dictionary invested more time investment on dictionary consultation than learners who used a click-on dictionary. However, on a subsequent unexpected spelling test the key-in group invested less time investment and performed better than the click-on group. The theoretical and practical implications of the results are discussed.

1. Introduction
In the digital age, the availability of authentic materials on the internet not only provides people with more opportunities to read digital text in unfamiliar languages, but also makes online reading an increasingly popular way of acquiring information (Abraham, 2008). Computer-mediated dictionaries have become important reading aids (Hulstijn, Hollander, & Greidanus, 1996; Laufer & Hill, 2000), and the ease of finding target vocabulary (i.e., headwords) is seen as its most important feature (Chen, 2010; De Ridder, 2002; Liu & Lin, 2011), as compared to paper-based dictionaries. Among computer-mediated dictionaries, the ease of finding headwords is different (Liu & Lin, 2011). The two most commonly used dictionaries are the key-in word dictionary and the click-on word dictionary. To look up an unknown word in a key-in dictionary, readers have to shift their attention from the text to the dictionary, type the unfamiliar word to look it up in the dictionary, and shift attention back to the text. To look up an unknown word in a click-on dictionary, readers have to click on the word with the mouse and shift their attention between the text and the dictionary too, but the attentional disengagement from the text takes less time. Previous research has shown that second language learners are more willing to look up words in a click-on dictionary than in a key-in dictionary (De Ridder, 2002; Liu & Lin, 2011). So, it seems that it is more convenient and time-saving to find a headword in a click-on word dictionary than in a key-in spelling dictionary.

Several second language (L2) studies have found that the dictionaries that learners use when they encounter unfamiliar words during reading can facilitate incidental vocabulary learning (Getzys, Imhof & Kautz, 2001; Hill & Laufer, 2003; Huckin & Coady, 1999; Hulstijn et al., 1996). However, it is not clear whether the use of dictionaries results in better reading comprehension (Liu & Lin, 2011), because reading comprehension involves cognitive processes other than word recognition, and studies on comprehension have used different
levels of instruction, text type, and types of assessment. In a meta-analysis of 11 studies that investigated the benefit of using computer-mediated glosses in second language reading comprehension and vocabulary learning, Abraham (2008) found a large effect on incidental vocabulary learning and an overall medium effect on second language reading comprehension. Particularly, a large effect size was found for multiple-choice reading comprehension tests which assessed learners’ receptive outcomes of comprehension. Most of the dictionary studies have focused exclusively on the acquisition of reading comprehension and word meaning skills. However, the incidental acquisition of word spelling knowledge through the use of digital dictionaries, which is also important in second language learning, has received much less research attention. In order to find the meaning of a word, learners must find the headword first. The task of finding the headword is a learner’s first step of noticing the form of an unfamiliar word, such as its spelling. In addition, this task contributes to vocabulary acquisition of word form. Once the form is familiar, figuring out what it means is the next step (Waring & Nation, 2004; Waring & Takaki, 2003).

While most learners find it more convenient to use a click-on dictionary than a key-in dictionary to look up words, an interesting question is how this difference in perceived convenience affects learners’ cognitive processes involved in looking up words and how it affects the incidental acquisition of vocabulary? Answering this question would not only allow us to get a better grasp of the effect of click-on and key-in target words on learners’ look-up cognitive processes and performances on vocabulary acquisition, but also allows us to ponder what comes with technological convenience.

Using computer-mediated dictionaries to assist second language reading can impose high working memory load, and from that point of view, it can be consider as a complex task. The complexity of this consultation task originates from the need to disengage attention from the text, shift it to the dictionary, and engage it on the dictionary when an unknown word is encountered and its meaning cannot be derived from the context. For the time visual attention is engaged on the dictionary, the disengaged text needs to be hold in WM, which imposes a high cognitive load. Although, it has been shown that looking up words in click-on dictionaries is experienced as more convenient than looking up words in key-in dictionaries (Liu & Lin, 2011), it is not clear how much cognitive load the consultation of these dictionaries will impose in the context of vocabulary acquisition. Paas, Renkl, and Sweller (2003) and Paas, Tuovinen, Tabbers, van Gerven (2003) have advocated the analytical identification of the interactive elements in a task (i.e. intrinsic cognitive load), the aspects of the task that interfere with schema construction and automation (i.e., extraneous cognitive load), the aspects of the task that are relevant or germane to learning, and the empirical determination of their cognitive consequences. Therefore, it is considered important to use cognitive load theory (Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merriënboer, & Paas, 1998) to identify the differential effects of the type of dictionary on vocabulary acquisition.

1.1. Cognitive load theory

Cognitive Load Theory (CLT) is used to develop instructional strategies that deal effectively with the limitations of our working memory (WM; Baddeley, 1986) and exploit the unlimited capacity of our long-term memory (Kalyuga, 2009; Sweller et al., 1998). Cognitive load can be defined as the load that a performance on a particular task imposes on the learner’s cognitive system (Paas, Tuovinen, et al., 2003; Paas & Van Merriënboer, 1994).

According to CLT, cognitive load can be productive (i.e., intrinsic load and germane resources) or unproductive (i.e., extraneous load) for learning (Paas, Renkl, et al., 2003; Paas, Renkl, & Sweller, 2004; Sweller, 2010; Sweller et al., 2011; Sweller & Chandler, 1994). Intrinsic cognitive load is generated by the materials to be learned and is imposed by “intrinsic nature of the material” (Sweller et al., 2011, p.57). Extrinsic cognitive load is generated by the instructional procedures that accompany the materials to be learned and is imposed by “the manner in which the material is presented” (Sweller et al., 2011, p.57). For learning tasks to be effective for learning, extraneous load should be as low as possible and as many WM resources should be devoted to information that is germane to learning, under the condition that the total cognitive load is within working memory limits (Paas, Renkl, et al., 2003; Paas, Tuovinen, et al., 2003).

According to Paas and van Merriënboer’s (1994) model, mental effort is the aspect of cognitive load that reflects the cognitive capacity that is actually allocated to accommodate the demands imposed by the task. To measure the amount of mental effort expended, subjective rating scales are the most commonly used empirical methods. Whereas, subjective rating techniques are very useful to determine the average load of a problem or the overall load across a series of problems, they are less useful to determine load at a more detailed level (Paas, Renkl, et al., 2003; Paas, Tuovinen, et al., 2003). Eye-tracking technology can be used to investigate cognitive load at a more detailed level by providing an online measure of visual attention allocation on a specific task, such as the time spent on the relevant reading tasks (Amadieu, Van Gog, Paas, Tricot, & Mariné, 2009; Van Gog & Scheiter, 2010).

1.2. Active processing and word recognition

One of the major differences between the two dictionaries is the way in which the headword is entered (click on target word vs. key in target word letters). Another major difference is the type of cognitive processes that is triggered in working memory by the headword finding task (passive storage spelling through clicking letter strings vs. active processing spelling through typing letters). Previous research provides important evidence that “learning by doing” can enhance deeper processing and visual letter recognition (James, 2010; Longcamp, Boucard, Gilhodes, & Velay, 2006; Vecchi, Richardson, & Cavallini, 2005). Particularly, visual-writing experience with the letters improved performance more than visual-only experience (James, 2010). Although Longcamp et al. (2006) found that handwriting characters contributes to visual recognition of letters more than typing them on a keyboard, in the present study, learners could not write letters to find headwords. However, all of the three experimental conditions required learners to visually read the words, and only two conditions further required learners to integrate the sensory and motor system. Therefore, we assumed that “reading the unknown word only” experience in the non-dictionary use condition would achieve lower performance on a spelling recognition test than in the two dictionary conditions. Furthermore, it can be assumed that typing in the correctly spelled unknown word in the key-in dictionary requires more involvement of the learners’ motor system in the spelling of the target word than just clicking on the correctly spelled unknown word in the click-on dictionary, which was expected to result in better performance on a spelling recognition test.
1.3. Vocabulary acquisition and dictionary consultation tasks

Efficient reading ability combines aspects of low-level processing (word recognition, syntactic parsing, and semantic-proposition encoding) and high-level processing (what text is about, how we decide to interpret the text, inferencing, background knowledge, etc.). Working memory (WM) supports word recognition, assembles the information at the word and clause level into a developing network of text-level information, and supports high-level text processing for comprehension (Grabe, 2009). Since learners have limited working memory capacity, computer-mediated dictionary can provide word knowledge to support word recognition, so it can free up some resources for managing other low-level and high-level cognitive processes while reading (Abraham, 2008; Grabe, 2009). To investigate the cognitive processes during dictionary consultation and their effects, we must specify which aspect of word knowledge learning (form, meaning, or use) (Nation, 2001) distinguishes click-on dictionaries from key-in dictionaries. Liu and Lin (2011) found that the use of a click-on or key-in dictionary enhanced the acquisition of word meaning (i.e., meaning knowledge), but did not differentially affect word meaning learning. However, they did not investigate the aspect of spelling learning (i.e., form knowledge).

Research on vocabulary acquisition has identified three factors that contribute to word retention: attention to word features, repeated exposure, and deep processing (Getty et al., 2001; Hill & Laufer, 2003; Koyama & Takeuchi, 2007; Krashen, 1989; Peters, 2007). Hulstijn (2001) emphasizes that the more a learner pays attention to a word’s different features and to intraword and interword relations, the more likely it is that the new word will be retained. However, researchers found that repeated exposures to a new word in language input are necessary to reinforce learning, but it is unclear how many repetitions are necessary (Huckin & Coady, 1999; Hulstijn et al., 1996; Laufer & Hill, 2000; Waring & Nation, 2004). In an analysis of the rate of learning depending on the frequency of occurrence, Waring and Takaki (2003) found that words that had been met more frequently were more likely to be learned. Furthermore, it is assumed that retention of information is determined by the way in which this frequency is processed (e.g., Laufer & Hill, 2000; Peters, 2007). To further analyze a dictionary consultation task, several subtasks, in which learners allocate their visual attention on characteristics of word form, can be identified to the acquisition of spelling knowledge. These subtasks include cognitive processes on: (1) paying attention to unfamiliar words when the word meanings cannot be extracted (i.e., attention to word features), (2) repeatedly looking at the unfamiliar words when the spelling cannot be remembered (i.e., repeated exposure and attention to word features), (3) taking actions to find headwords when learners decide to look up words in the dictionaries (i.e., deep processing), or paying attention to word syllables in the dictionary content (i.e., attention to word features). Performing these dictionary consultation tasks enable learners to deeply process language input and develop language awareness.

1.4. Eye tracking

Besides using rating scales to collect learners’ subjective cognitive load, many researchers have used log files to study cognitive processes (e.g., Chen, 2010; Liu & Lin, 2011). However, the log file design does not allow a further division of the actual time spent on the computer-mediated dictionary definition window and on reading text. It is even impossible to find where learners actually have looked on the screen. Over the past decades, researchers have been using eye movement behavior to infer cognitive processes (Rayner, 1998; Rayner & Duffy, 1986; Simonsen, 2011; Tono, 2011). Based on Just and Carpenter’s (1980) eye-mind assumption, the eyes remain fixated on a word as long as the word is being processed, so the gaze time was used to reflect how much cognitive effort the learners put into a specific task. Therefore, the cognitive processes on the consultation subtasks described above can be represented through the following gaze behaviors: (1) attention to word features is measured through fixation duration on target words and, (2) repeated exposure is measured through fixation frequency on target words, (3) deep processing is measured through headword finding duration and total fixation duration and time investment on spelling relevant tasks. The gaze behaviors that learners allocate to these eye-tracking measures can be considered as cognitive resources that are effective for learning.

Only a few studies have applied eye-tracking technology to analyze the process of dictionary look-up. Simonsen (2011) and Tono (2011) were pioneers who applied eye tracking technology on dictionary studies. However, their research focus was not the same as in the present study, because they did not investigated the reading of passages with dictionary consultation using eye-tracking technology. In Tono’s (2011) study, one cue sentence was presented to learners once, and learners did not have to search for the dictionary entry. In reality, learners have to find a headword entry themselves. On the other hand, Simonsen (2011) was interested in production of a translation task, not in reading. Therefore, there are no studies that have investigated the concurrent processes of reading passages with dictionary consultation using eye-tracking technology.

To further explore learners’ gaze behaviors during vocabulary acquisition and reading comprehension, the present study applied eye-tracking technology to record and analyze learners’ dictionary look-up processes (Just & Carpenter, 1980). In addition, the present study focused on both word spelling learning and reading comprehension performances. Two experimental groups including a key-in word dictionary group and a click-on word dictionary group were compared to discover where and how much cognitive load is imposed on the learners’ cognitive system during the dictionary look-up process. In addition, both experimental conditions were compared to a non-dictionary control condition.

1.5. Hypotheses

Based on the results of the studies on second language reading of De Ridder (2002) and Liu and Lin (2011) we hypothesized that participants who use a click-on dictionary would be more willing to look up words, and consequently consult the dictionary more frequently than participants who use a key-in dictionary (Hypothesis 1). The main research question that we explore in this study is which dictionary format will lead to better learning of word spelling. The process of looking up words in a click-on dictionary, in which the learner can simply click on the unknown words with the mouse, is expected to require less mental effort and cognitive resources (i.e., gaze time investment) than looking up words in a key-in dictionary, in which the unknown words need to be typed on the keyboard. However, it is believed that the higher effort for actively typing letters in the key-in dictionary is effective to the learning of word spelling (i.e., impose higher germane resource), because the typing activity focuses the
learner’s attention more closely and actively on the spelling of that word (James, 2010; Longcamp et al., 2006; Vecchi et al., 2005). Based on the positive results from sensori-motor experiences with word recognition, in the present study, all of the three experimental conditions required learners to visually read the words, and only two conditions further required learners to integrate sensory and motor systems. Therefore, it is hypothesized that consultation of a key-in dictionary will impose a higher total cognitive load (i.e., longer fixation duration on target words) to actively type spelling than consultation of a click-on dictionary (Hypothesis 2), and that both dictionary groups will impose higher total cognitive load (i.e., longer fixation duration on target words) than a non-dictionary control group (Hypothesis 3). According to Van Gog and Paas (2008), if learners invest more mental effort in relevant processes in the learning phase, they will construct better cognitive schemata, which will enable them to achieve higher performance with less mental effort investment on a subsequent test. Therefore, it is hypothesized that the higher load in the key-in condition duration the intervention stage will achieve higher performance (Hypothesis 4a) and experience less cognitive load (4b) on an unexpected spelling test than learners who used a click-on dictionary. In addition, the key-in dictionary group is expected to achieve higher performance (Hypothesis 5a) and experience less cognitive load (Hypothesis 5b) on the spelling test than the non-dictionary control group. It is not clear what the difference between the click-on group and control group will be.

A computer-mediated dictionary can provide word knowledge to support word recognition, so it can free up some resources for managing other low-level and high-level cognitive processes while reading (Abraham, 2008; Grabe, 2009). Therefore, it is hypothesized that second language learners who used a dictionary will achieve higher reading comprehension performance on multiple-choice items than learners without a dictionary aid (Hypothesis 6).

2. Method

2.1. Participants

Forty-five undergraduate students from a Taiwanese university, who were aged between 18 and 23 years ($M = 20.4$ years, $SD = 1.26$ years; 23 females and 22 males) with different majors participated in the study. Their native language was mandarin Chinese and they learned English as their foreign language as a part of their first year course. All participants had passed an eye-tracking calibration and validation procedure.

2.2. Design

The treatment (independent variable) used in the experiment includes three conditions: (1) click-on dictionary (Click-on group, C); (2) key-in dictionary (Key-in group, K); and (3) non-dictionary (Control group, N). Participants were randomly assigned to one of the three experimental groups, in such a way that each condition contained 15 participants. In the click-on condition, learners were instructed to click...
once with the mouse on a target word, after which the headword and its definition were presented (Fig. 1a). In the key-in condition, learners were instructed to click once on the headword finding field to bring forward a small screen keyboard (Fig. 1b), and then key in the spelling, which would bring forward the definition content (Fig. 1c). In the no dictionary condition, learners were instructed to read materials without dictionary aids (Fig. 1d).

English proficiency, reading material, dictionary material, and computerized environment were either controlled or standardized in both conditions. First of all, before the experiment, all participants took the English proficiency tests to evaluate participants’ vocabulary and reading comprehension expertise after reading an article. This pretest included 10 multiple-choice spelling questions, 10 matched form-meaning questions, and 9 multiple-choice reading comprehension questions. Descriptive information is summarized in Table 1. Analysis of variance (ANOVA) of participants’ English proficiency and spelling ability scores indicated that there were no differences across conditions at the start of the experiment, $F(2, 42) = .320, p = .728$, and $F(2, 42) = .021, p = .979$, respectively.

The reading material, Birth-Order Myth, was written to meet the undergraduate students’ interest and English proficiency. It was written by an English professor who is a native-English speaker. The reading material consisted of four paragraphs with “compare and contrast” text structure and 475 words. Using the Frequency Level Checker (Maeda, 2010), 6.67% of the words were identified as words found most frequently in secondary and university level reading material, but did not include any words from Levels 1 and 2, and 11.88% of the words was outside the three frequency levels mentioned above. The reading material was reviewed by two EFL teachers and 10 college students, who were not involved in the experiment. The feedback indicated that the reading content was interesting and the topic suitable for college students. In addition, the content of the dictionaries was from the Oxford Advanced Learner’s English-Chinese Dictionary (7th edition). All 475 words in the reading text could be found in both computer-mediated dictionary conditions, and participants were free to check any words they wished.

Finally, all materials and tests were computerized. To facilitate eye tracking specific guidelines were used for the design of the materials. First of all, the reading passages were displayed in the left column, and dictionary content in the right column. In this way, participants could move their eyes left to right and right to left to search information. Second, instead of using the scrollbar to navigate pages, previous/next page buttons were created to facilitate the analyses of the eye tracking data. Third, each page was labeled as “current page/total page”, so participants could track pages to search information. Furthermore, a small screen keyboard was created for the key-in dictionary group, which enabled participants to maintain their heads in an upright position during the task. In this way, the eye tracker could keep track of the participants’ eyes all the time. Furthermore, learners in both click-on and key-in groups look up the same amount of information.

Planned comparisons and effect size indices $d$ were used to detect any significant differences between (1) the key-in dictionary group and the click-on dictionary group, (2) the key-in dictionary group and the control dictionary group, and (3) the average dictionary groups and non-dictionary control group. The level of significance was set at $p < .05$. According to Cohen’s (1988) rough characterization, $d = .2$ is deemed to be a small effect size, $d = .5$, a medium effect size, and $d = .8$ a large effect size.

2.3. Instruments

To evaluate the participants’ acquisition of spelling knowledge and reading comprehension, an immediate multiple-choice posttest was designed. Two types of cognitive load measures were used to measure learners’ mental effort and time investment. One method was to ask the learner to provide a subjective rating on a questionnaire. The other method was based on the learner’s eye movement indices through eye-tracking technology.

2.3.1. Spelling and reading comprehension multiple-choice posttests

An unexpected spelling multiple-choice posttest was designed to evaluate participants’ spelling learning performance. In the evaluation stage of the reading material, prior to the experiment, ten students were asked to mark unfamiliar words while reading the passage. After considering word familiarity and relevancy, 15 unfamiliar but relevant target words were identified and were tested for spelling. Target words included endowment, primogeniture, agape, squabbling, siblings, chromosomes, subordinate, manifest, diluted, prophecy, ascribed, superfluous, cognitive, hierarchical, and egalitarian. The multiple-choice distracters were chosen from commonly found spelling errors of Taiwanese students (Lin, Huang, Chen, & Hou, 2009). Participants had to choose the correctly spelled word from four options in the spelling multiple-choice posttest. The reading comprehension test consisted of 12 reading comprehension multiple-choice questions in English without time restrictions, which measured participants’ basic comprehension, inference, and summary abilities. After the experimental materials and tests were designed, they were sent to EFL college instructors for expert construct review. After a first run of reliability and item analyses, both the passage and test items were reviewed and revised. ITEMAN 4.1 was used to analyze test-level reliability (Coefficient α) and item-level difficulty ($P_{value}$) and discrimination (Person point–biserial correlation, $r_{pbis}$). The reliability index of the spelling test and reading comprehension test was .881 and .752, respectively. These indices were based on 100 students who participated in the test construct and reliability analyses.

2.3.2. Mental effort: rating scale

Participants reported their invested mental effort in the unexpected spelling test on Paas’s (1992; Paas, Van Merriënboer, & Adam, 1994) 9-grade symmetrical category scale, by translating the perceived amount of mental effort into a numerical value. The numerical values and

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means and standard deviations (in parentheses) for English proficiency.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control M(SD)</th>
<th>Click-on M(SD)</th>
<th>Key-in M(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English proficiency test score (range 0–29)</td>
<td>17.27 (3.43)</td>
<td>16.87 (3.99)</td>
<td>17.87 (2.80)</td>
</tr>
<tr>
<td>Spelling performance (range 0–10)</td>
<td>7.07 (2.15)</td>
<td>7.20 (1.57)</td>
<td>7.13 (1.59)</td>
</tr>
<tr>
<td>Word meaning performance (range 0–10)</td>
<td>4.53 (2.59)</td>
<td>4.87 (2.20)</td>
<td>5.07 (1.53)</td>
</tr>
<tr>
<td>Reading comprehension performance (range 0–9)</td>
<td>5.67 (1.63)</td>
<td>4.80 (1.37)</td>
<td>5.67 (1.05)</td>
</tr>
</tbody>
</table>
labels assigned to the categories ranged from very, very low mental effort (1) to very very high mental effort (9). The participants were asked How much mental effort did you invest in the spelling test?

2.3.3. Mental effort: eye-tracking technology

To record participants’ eye movement data, an EyeLink 1000/2k Eye Tracker was used. The eye tracker was discretely integrated into a 19” TFT-LCD monitor in the ratio of 5:4 (1280 × 1024 pixels). A 35 mm camera lens and an adjustable forehead-chin rest were used. The participants’ viewing distance was fixed at 70 cm. Participant’s viewing was binocular, but eye movements were recorded from the right-eye only at a sampling rate of 250 Hz. GazeTracker 9.0 eye movement data recording and analysis software was used to manage and analyze the data. Fixations were identified as a set of gaze points that fell within a 40-pixel dispersion and together lasted for at least 200 ms (Reichle, Pollatsek, Fisher, & Rayner, 1998). The length of the paragraph was chosen in such a way that participants did not have to make vertical head movements. The font size was 14-point with Arial font style. To determine the perfect line spacing, 6 participants were tested for three experimental groups in the pilot study, and the line space was finally set to 1.95.

Based on Just and Carpenter’s (1980) eye-mind assumption, gaze times on target words were used to reflect how much mental effort the learners invested in a specific task. To test the hypotheses, one AOI (Area of Interest) in the reading text field (AOI-T) and two AOIs in the dictionary field (AOI-H, AOI-S) were setup. Fig. 2 shows a target word (endowment) example:

1. **AOI-T**: this area contains the target word only, such as AOI-T (endowment).
2. **AOI-H**: this area contains the headword finding field, such as AOI-H (查詢單字: Endowment) (Fig. 2a). Learners in the Key-in group could use a small screen keyboard to key in word spelling. Learners in the Click-on group could click once on word in the reading text field, after which the clicked word would appear in the headword finding field (Fig. 2b).
3. **AOI-S**: this area contains the target word separated by syllables, such as AOI-S (en.dow.ment).

![Fig. 2. Treatment conditions with AOI types and locations. (a) Key-in dictionary with AOIs. (b) Click-on dictionary with AOIs.](image-url)
2.4. Variables

Eleven dependent variables related to frequency of dictionary usage, time investment on AOIs, spelling and reading comprehension performances, and cognitive load measurement were defined. The following two indicators were used to investigate frequency of dictionary usage:

1. **Consultation frequency of all text words** was represented by total number of consulted word on the reading materials, not restricted to target words.

2. **Consultation frequency of target words** was represented by total number of consulted target words in the dictionaries.

In addition, four cognitive processes and eye movement indices were proposed to test time investment and vocabulary acquisition:

3. **Total gaze duration of target word** was calculated when a target word was fixed more than one fixation; which is the total fixation duration on AOI-T. Furthermore, total gaze duration of target word was separated into four stages: before looking it up in the dictionary, during finding a headword in the headword finding field, during reading the dictionary content, and after looking it up in the dictionary. This variable was used to investigate the degree of attention paid to the word features.

4. **Frequency of target word fixation** was the total number of the same target word fixated. It was used to investigate the frequency of repeated exposure to the word features.
(5). **Headword finding duration** was calculated as the time a learner invests on finding a headword. For learners in Key-in group, the headword finding duration indicated the time starting from a small keyboard was opened to the small keyboard was closed until the correct headword was displayed on AOI-H. The headword finding duration also included the total gaze duration of the target word during looking it up in the Key-in dictionary. For learners in Click-on group, it was the mouse click time. This variable was the major difference between click-on and key-in dictionaries.

(6). **Total fixation duration and time investment on spelling relevant tasks** is the sum of total gaze durations on the target word (AOI-T) before looking it up and after looking it up in the dictionary, headword finding duration, headword fixation duration (AOI-H) after finding the headword, and syllable fixation duration (AOI-S). This gaze time variable was used to indicate mental effort (i.e., fixation duration) on spelling relevant tasks based on the Eye-Mind Assumption of Just and Carpenter (1980).

The following four indicators were used to evaluate dictionary consultation usage and performances of spelling and reading comprehension:

(7). **Spelling performance** was represented by the learners’ immediate posttest spelling scores.

(8). **Reading comprehension performance** was represented by the learners’ immediate posttest reading comprehension scores.

(9). **Accuracy of spelling learning on consulted items** was represented by the performance on consulted items divided by the total number of consulted items. The unit of measurement will be in percentage to be able to compare it between the two dictionary conditions. It is used to investigate whether looking up target words in the dictionary affected the accuracy of spelling outcomes, spelling performance on consulted target words was compared.

(10). **Accuracy of spelling learning on non-consulted items** was represented by the performance on non-consulted items divided by the total number of non-consulted items. It will be in percentage and be able to compare it between two dictionary conditions.

The following one indicator was used to investigate subjective mental effort after taking the posttest.

(11). **Mental effort invested in the spelling test** was represented by the subjective rating scale.

2.5. **Procedure**

Participants were tested in individual sessions in three stages. The procedure for the study is shown in Fig. 3. At the first stage, participants filled out a consent form, and were seated in front of the monitor and the eye tracking system, with the head fixated on the chin rest at a distance of 70 cm and 60 cm between the eyes and the stimulus presentation monitor and the eye tracker, respectively. The first time a 13-point calibration and validation procedure was performed to make sure that participants’ eye condition was suited for joining in the experiment. After filling out personal information, the participants received a pretest on their English proficiency without time limitation. After finishing the pretest, they left the lab and came back for the experiment within 2 weeks.

At the second stage, each participant was randomly assigned into one of the three experimental groups, and an experimental tutorial was provided to make sure that participants were familiarized with experimental tools. They were told that their reading comprehension performance would be tested. Next, a 13-point calibration and validation procedure of the eye-tracking system was performed. Screen contents, mouse click actions, and eye movements were recorded.

At the last stage, participants were administered an unannounced immediate spelling and reading comprehension posttests and cognitive load questions. Finally, participants were debriefed and thanked for their participation.

3. **Results**

In this section, five results are presented in the following order: (1) frequency of dictionary usage, (2) eye-tracking measures of mental effort allocation on the dictionaries, (3) performance on spelling knowledge, (4) subjective ratings of mental effort invested in the spelling test, and (5) performance on reading comprehension. Planned comparisons were used to test the specific hypotheses about the differences between (1) the key-in dictionary group and the click-on dictionary group (Hypotheses 1, 2, 4a, 4b), (2) the key-in dictionary group and the non-dictionary control group (Hypotheses 5a and 5b), and (3) the average of the dictionary groups and the non-dictionary control group (Hypotheses 3 and 6).

3.1. **Frequency of dictionary usage**

The Click-on group ($M = 47.60, SD = 21.32$) consulted the dictionary significantly more often to look up unknown words than the Key-in group ($M = 30.07, SD = 16.64$), $t(28) = 2.511, p = .009$, one-tailed, $d = .92$, which confirmed Hypothesis 1. However, there were no differences between the Click-on group ($M = 13.20, SD = 1.86$) and the Key-in group ($M = 11.53, SD = 4.32$) in the frequency to consult the dictionary for the 15 target words, $t(28) = 1.371, p = .093$, one-tailed, $d = .50$.

3.2. **Eye-tracking measures of mental effort allocation on the dictionaries**

To further investigate learners’ mental effort allocation, eye-tracking technology was used to discover where and how learners devote their cognitive resources on dictionary consultation. Several eye movement indices were collected and summarized in Table 2. In terms of total gaze duration of target words, the results indicated that the Key-in group spent significantly more time paying attention to the target word features than the Click-on group, $t(28) = 4.050, p < .001$, one-tailed, $d = 1.48$. Furthermore, for the frequency of target word fixation,
the results indicated that the Key-in group fixated more often on the target word features than the Click-on group, t (28) = 2.943, p = .003, one-tailed, d = 1.07.

Spelling relevant tasks include the sum of total gaze duration on the target word (AOI-T fixation) before and after looking it up in the dictionary, headword fixation duration (AOI-F finding), headword fixation duration (AOI-H fixation), and syllable fixation duration (AOI-S fixation). The results indicated that the Key-in group spent significantly more time on total fixation duration and time investment on spelling relevant tasks than the Click-on group, t (28) = 8.916, p < .001, one-tailed, d = 3.26. These results confirmed Hypothesis 2 that consultation of a key-in dictionary imposed higher cognitive load than consultation of a click-on dictionary.

Furthermore, the results of the analysis on the total gaze duration on target words, t (43) = 3.138, p = .001, one-tailed, d = 1.08, the frequency of target word fixation, t (43) = 3.556, p < .001, one-tailed, d = 1.24, and total fixation duration and time investment on spelling relevant tasks, t (43) = 3.564, p < .001, one-tailed, d = 1.30, confirmed Hypothesis 3 that both dictionary groups imposed higher cognitive load than a non-dictionary control group based.

3.3. Performance on spelling knowledge

The descriptive statistical analyses of the different performance measures for the two dictionary groups are summarized in Table 3. For spelling performance, the result confirmed Hypothesis 4a. A one-way ANCOVA was performed after confirming the requirement of homogeneity within-cell regressions (F(1,26) = .559, p > .05), with the pretest scores of English proficiency as a covariate. After adjusting for the pretest scores, there was a significant difference between the two dictionary groups regarding spelling performances (F(1,27) = 5.529, p < .05, d = 3.34). The adjusted mean scores of spelling performances was higher in the Key-in group (11.64) than in the Click group (9.63).

Furthermore, the results confirmed Hypothesis 5a. A one-way ANCOVA was performed after confirming the requirement of homogeneity within-cell regressions (F(1,26) = .758, p > .05) and adjusting for the pretest scores, there was a significant difference between the Key-in group and Control group regarding their spelling performances (F(1,27) = 6.241, p < .05, d = 3.54). The adjusted mean scores of spelling performances was higher in the Key-in group (11.61) than in the Control group (9.32).

To further investigate whether looking up target words in the dictionary affected the accuracy of spelling outcomes, spelling performance on consulted target words was compared. For the accuracy of spelling learning on consulted items, the results indicated that the Key-in group had a significantly higher possibility on giving a correct spelling answer when they consulted target words than the Click-on group, t (28) = 2.753, p = .005, one-tailed, d = 1.00. For non-consulted target words, the results indicated that no difference was found between Click-on and Key-in conditions, t (28) = .624, p = .269, one-tailed, d = .23.

3.4. Subjective ratings of mental effort invested in the spelling test

For the subjective mental effort ratings between the Key-in group and the Click-on group, the results confirmed Hypothesis 4b that readers on Key-in condition invested significant less cognitive load on taking spelling test than learners on Click-on condition, t (28) = 2.172,
p = .019, one-tailed, d = .80. On the other hand, the difference between the Key-in group and non-dictionary control group was not significant, t (43) = 1.661, p = .054, one-tailed, d = .61. Therefore, Hypothesis 5b was not confirmed.

3.5. Performance on reading comprehension

For reading comprehension test performance, a one-way ANCOVA was performed after confirming the requirement of homogeneity within-cell regressions (F(1,41) = .774, p > .05), with the pretest scores of English proficiency as a covariate. After adjusting for the pretest scores, there was no significant difference between the both dictionary group and Control group regarding reading comprehension performances (F(1,42) = .016, p > .05, d = .18). The adjusted mean scores of reading comprehension performances was higher in the both dictionary group (7.43) than in the Control group (7.33). Therefore, Hypothesis 6 could not be confirmed.

4. Discussion

The main purpose of this study was to investigate whether digital dictionary format differentially affects students’ incidental acquisition of spelling knowledge and cognitive load during second language learning. The main hypotheses that using a key-in dictionary is more effortful but leads to less effortful and higher spelling performance than using a click-on dictionary were confirmed. The combination of mental effort data that were collected with the eye tracker during the use of the dictionary and the subjective rating scale during the posttests, and the performance on the unexpected spelling test clearly showed that the time investment and sensori-motor experiences that were involved in looking up words in the key-in dictionary was effective to the incidental acquisition of spelling knowledge.

Based on vocabulary acquisition hypotheses (attention to word features, repeated exposure, deep processing), several look-up cognitive processes associated with gaze behaviors on dictionary consultation activities were identified as effective to the incidental acquisition of spelling knowledge. Based on the concept of active learning, the integration of visual reading and motor typing spelling in key-in dictionary condition should contribute to better spelling recognition performance. Since the major difference between click-on and key-in dictionary is the level of cognitive load required for finding headwords in the dictionary through actively entering the letters in the key-in dictionary, did learners who invest mental effort on finding headwords and spelling relevant tasks contribute to better spelling performances?

Because of the less mental effort and the involvement of motor actions in finding headwords, we hypothesized and confirmed Hypothesis 1 that participants were more willing to use the click-on dictionary to look words up than key-in dictionary. However, the results confirmed that the increased clicking behavior did not increase the spelling learning performance (Hypotheses 4a). If the frequency of usage did not affect the spelling learning, what factor contributed to higher spelling performance on key-in condition? From different perspective, we suspected that the higher spelling performance on key-in condition resulted from consultation task itself. The results of spelling test performance, shown in Table 3, confirmed our assumption that when target words were consulted, the possibility of getting a correct spelling answer was significantly higher in the key-in condition. Therefore, we further investigated the consultation process by analyzing mental effort allocation in the two dictionary conditions and the non-dictionary condition.

Although the participants in the key-in group experienced a higher cognitive load from dictionary consultation than learners in the click-on group (Hypothesis 2), as indicated by the four eye-tracking indices (shown in Table 2), they experienced less cognitive load from an unexpected spelling test (Hypothesis 4b) and achieved higher spelling performance in the posttest (Hypothesis 4a). With regard to Hypothesis 4, this combination of mental effort and performance data showed that the mental effort that was invested in looking up words through the integration of visual reading and motor typing spelling in the key-in dictionary was effective to the incidental acquisition of spelling knowledge. Furthermore, all of the three experimental conditions required learners to visually read the words, and only two dictionary conditions further required learners to integrate sensory and motor systems. Both dictionary groups experienced a higher cognitive load than the non-dictionary control group (Hypothesis 3). The key-in dictionary group performed significantly better on spelling performance than the click-on dictionary group (Hypothesis 4a) and the no-dictionary group (Hypothesis 5a). In addition, the key-in condition experienced less cognitive load on the spelling test than click-on condition (Hypotheses 4b). Besides, although the result of testing Hypothesis 5b did not confirm our assumption that the key-in dictionary group experienced less cognitive load on taking the spelling test than the non-dictionary control group, the results were in the expected direction and approached significance (p = .054).

For reading comprehension performance, the results of testing Hypothesis 6, shown in Table 3, did not confirm our assumption that learners used dictionaries (key-in and click-on) achieved higher reading comprehension performances than learners without dictionary aid. Liu and Lin (2011) found the same result that dictionary consultation did not facilitate reading comprehension. The typing of the word during the dictionary look-up process may improve incidental vocabulary and spelling acquisition, reading comprehension does not seem to profit from it. Although the computer-mediated dictionary provided word knowledge to support word recognition (Abraham, 2008; Grabe, 2009), it seems that inferring the meaning of an unknown word from the context is just as good for reading comprehension than looking it up in a dictionary. A possible reason for this finding is that looking up an unknown word in a digital dictionary disengages readers’ attention from the text, which leads to a split-attention effect (e.g., Ayres & Sweller, 2005; Bobis, Sweller, & Cooper, 1993; Kalyuga, Chandler, & Sweller, 1999). For future research it would be interesting to investigate this split-attention hypothesis by comparing a dictionary format in which the unknown word is replaced by the dictionary result at the same physical location, to the conventional digital dictionaries and a non-dictionary control condition. The physical location should not cover the contextual words around unfamiliar words as the pop-up window within the text that interfere learners read the contextual sentences for comprehension.

The significance of the present study was the further investigation of the cognitive process of dictionary consultation and spelling performance, in contrast to Liu and Lin’s (2011) study, in which word meaning performance was studied. The results confirmed the benefit of typing spelling into the target headword finding field, so learners had the opportunity to notice and pick up the spelling of the unknown words through integrating sensory and motor systems (James, 2010; Longcamp et al., 2006). The current study, however, provides an important implication to computer-mediated dictionary users and language teachers: Stimulating second language learners’ active involvement in finding the headword facilitated form acquisition. As mentioned, vocabulary learning can be enhanced by task demands (Huckin & Coady, 1999), the headword finding task induced spelling relevant tasks, and both types of tasks enhanced spelling acquisition. While many learners use the more convenient click-on dictionary to look up words in the dictionary, learners have to be aware that they lose
the opportunities to pay more attention to word features, repeatedly exposure to word feathers, and perform deep processing of language input. More importantly, they lose the opportunity to retain the word knowledge although they may save some time and mental effort on dictionary consultation. As Hulstijn and Coady (1999) noted, if vocabulary learning requires a precise and effortful coordination of form and meaning, it may not optimally occur with an activity that allows the learners to bypass such precision and effort, such as using click-on dictionary to bypass typing spelling.

The results of this study revealed some areas that are needed for further investigation. First, although results of pretest indicated that both dictionary groups had approximately the same spelling pretest performance, the general vocabulary size of these two groups may be different (Nation, 2001). The vocabulary size may be reflected in the frequency of consultation. People who have a larger vocabulary size may not need to consult a dictionary as frequently as people who have a smaller vocabulary size. Second, the current study did not look at the qualitative aspect of keying spelling: Do learners memorize the whole spelling before keying spelling into the headword searching field? This question needs to be answered in future research. Furthermore, the participants in this study were Taiwanese with Mandarin Chinese as their first language, which is a non-alphabetic language. They might not be used to the English orthography. Future studies could investigate the orthography influences on spelling acquisition through comparing non-alphabetic first language learners with alphabetic first language learners (e.g., Spanish). In addition, the relationship between learners’ individual differences and dictionary consultation behaviors is essential to provide pedagogical strategies for learners and scholars. Further study can investigate the consultation behaviors between high and low reading proficiency participants. Finally, the specific design of the dictionaries’ interfaces might have affected the results. It could, for example, be argued that a design in which the spelling of the unknown and looked-up words would have been presented in close proximity to the unknown word, could lead to different effects on incidental spelling acquisition and second language learning. Future research could investigate the influence of the interface design.

Acknowledgements

The authors would like to thank the kind assistance and helpful comments of the editor of Computers & Education and the anonymous reviewers of this paper. We also would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 100-2628-S-008-001-MY3, NSC 101-2511-S-008-004-MY3, NSC 101-2811-S-008-004, and NSC 102-2811-S-008-014. Furthermore, we would like to thank Oxford University Press for granting permission to use Oxford dictionary content in this study. Finally, we would like to thank the administration, the teachers and the students of National Central University who participated in this study.

References


