Investigating Using Behaviors of E-dictionary with Multiple Design: A Perspective from the Integration of Eye-Tracking Technique and Stimulated Recall

Xuesong Zhai
School of Educational Technology
Beijing Normal University
Beijing, China
zhxs@mail.ustc.edu.cn

Nanxi Meng
Department of Learning Technologies
University of North Texas
Denton, TX, USA
nanxi.meng@unt.edu

***Abstract*—Electronic Dictionary (ED) was one of the most popular learning tools with its increasingly diversified design ranging from information searching to knowledge management. Although previous research has surveyed in terms of the relation between ED and learning outcomes, few studies employed objective measurements, such as biofeedback, along with a qualitative approach to investigate learners' using behaviors and their cognitive process. The current study proposed to capture participants' three eye movement dates including fixation counts, fixation duration and scanning path, which were utilized as the stimulus to help participants to retrospect their cognitive features. The fixation interestingly shows that the areas of interest are somewhat impaired with their expectation in pre-survey, for example, users claimed they prefer as many as example sentences, while only the first three were noticed. Moreover, the scanning paths frequently moved back and forth among some elements, which indicate learners tried to make connections among some specific elements to help enhance learning efficiency. The stimulated recall interview was employed to further explore their cognitive process based on their eye movement data, and some implications were made for the optimization of the ED design to meet learners' personalized requirements.**

Keywords—eye tracking, e-dictionary, stimulated recall, cognitive behavior

I. INTRODUCTION

Electronic dictionary (ED), featuring sufficient contents and ease of use, has been considered as efficient supplement and support for information searching [1] and knowledge management in the digital age [2], which are increasingly attracting abundant customers. To meet users' individualized requirement, some researchers and designers have conducted the surveys on users' perceptions of the ED in terms of the contents and interface design [3]. Although users' self-reported responses were referable to the improvement and re-designing of EDs, there still existed several challenges. Typically, the self-reported survey has such subjective nature of introspection that the objective measurements are warranted to present instant using behavior. Besides, self-reported survey merely provide users' overview perceptions of EDs, and fail to investigate users' personalized using behavior and understanding of EDs [4], which warrant the further research.

To fill this research gap, the eye-tracking technology was proposed to be employed to capture the participants' eye movement data, including fixation count, fixation duration and scanning paths, which visualize the users' eye movements regulation and their basic cognitive processes through heat maps and scanning figures as using EDs. Based on the Cognitive Theory of Multimedia Learning and the Eye-mind assumption, it is suggested that users' eyes movements could accurately reflect how they process information from the interface of EDs, as well as their cognitive behaviors, which provide practical suggestions to improve the design of EDs and facilitate the individualized learning. Moreover, in order to get access to the deep understanding of these eye-tracking data, the Stimulated Recall Interview (SRI) are employed to examine users' retrospection of their substantial usage and cognitive process, with the stimulus of the heat maps and scanning figures.

Above all, this study aims to explore the following research questions:

1. What were cognitive behavior presented when they viewed the ED pages?
2. How should the design of EDs be improved to meet users' individualized learning requirement?

II. THEORETICAL FOUNDATIONS AND RESEARCH BACKGROUND

A. The Cognitive Theory of Multimedia Learning

The Cognitive Theory of Multimedia Learning (CTML) by Mayer is based on Dual coding theory, working memory theory, Constructivism and Cognitive Load Theory (CLT, [5][6]. Applying the types of cognitive load in multimedia learning environment, CTML reveals the cognitive process when study with multimedia learning tools.

CTML consists three assumptions about people learn with multimedia: 1) dual coding: Human information processing system includes dual channels for visual/pictorial and auditory/verbal processing, separately. The designing of

978-1-5386-6522-0/18/$31.00 ©2018 IEEE

multimedia that facilitate building connections between pictorial and verbal representations could assist in meaningful learning, evidence could be seen in vocabulary learning [7]. 2) limited capacity: each channel has a limited capacity for processing. So far information and data overload were identified in many online dictionaries, which distract users from obtaining information efficiently. Online dictionary with low cognitive load provided best learning performance [8]. 3) generative learning: generative learning happens when learners carry out a coordinated set of cognitive processes during learning, these processes include essential processing, extraneous processing and generative processing [9].

Learners' cognitive processing, which usually relies on the complexity of material (i.e., the number of interacting elements), is essential for comprehending the presented material. The design of interactive elements like hyperlinks, glossary book and voting on ED pages should promote cognitive processing but avoid increasing the complexity of for users. Extraneous processing happens when learner engages in cognitive processing that does not support the learning objectives. To better facilitate information seeking, the pictures, examples and texts on ED pages should be designed interactively and present in favor of viewing. To achieve generative processing, learner needs to engage in deep cognitive processing such as mentally organizing the material and connect to prior knowledge, which requires EDs provide daily life-oriented materials users could associate with.

To sum up, in pursuance of receiving generative processing for ED users, EDs should be necessarily designed and presented according to human cognitive architecture, take learner's cognitive load into consideration and efficiently contribute to information capture and understanding [10]. Based on the integration of the existing multimedia information and users' previous experiences, users are capable of constructing their accordant psychological representations. These representations could facilitate to memorize and improve learning efficiency [11].

B. Eye tracking technique

Eye tracking technique provides a direct channel to gain information about human behavior by observing, measuring and interpreting human behavior [12]. Eye tracking approach is developed based on the eye movements and the eye-mind link [13], which indicates that via eye movements, a dynamic trace of attention can be observed, hence the degree of cognition plays an active role in guiding the eye during reading can be reflected. It is assumed to represent the perceptual and cognitive processing of stimuli, during which the eyes remain relatively concentrated, allowing acquiring information from the part of the stimulus where attention was paid.

Eye movements consist of three basic evaluative criteria: fixation counts, fixation duration, and scanning paths [14]. Firstly, fixation counts are the fixations counted in an area of interest (AOI) or a task. On average, a fixation count was defined last between 200 and 300 milliseconds. Fixation counts are usually used to identify the part of a screen or material that is viewed, reveal patterns describing how a user's attention is directed to a given region or visual area of the computer screen. Moreover, it could be used to gauge the complexity level of the image and the sequence of viewing [15]. Secondly, fixation duration refers to the fixation duration time within a word or an AOI; it is the eye movement measured on the time scale. Fixation duration is affected by factors like the type of activity performed, or the complexity of the materials. Thirdly, scanning path reveals the order or sequence of the fixations, helping to depict the experience a user has while engaging with viewing materials [16]. Scanning path is adopted to get access to visual memories. When a particular sequence of eye movement is executed, the pattern is viewed, and visual memory for the pattern is reflected.

Eye tracking is an adaptable approach to investigate the cognitive processes and mechanisms involved in the information acquiring and multimedia designing. The previous study has utilized eye is tracking to examine the complexity of webpages from the perspective cognitive load, revealing the correlation between complexity of webpages, user's attention and their cognitive load [17]. In the recent educational research, the eye-tracking technique is popularly employed to explore learning processes in complex learning contexts in multimedia learning [18][19]. Employing interfaces of multimedia tools as stimuli, eye-tacking has been regarded as a supplement method to investigate cognitive processes linked to learning and performance [20].

C. Stimulated Recall Interview

The data collected from Eye-tracking techniques need to be further explained in terms of why the AOI was paid more attention, which need learners' review and retrospection. Stimulated Recall Interview (SRI) as an introspection procedure, has been used extensively in educational research in learning [21]. The recorded multimedia are offered to learners to stimulate recall of their concurrent cognitive activity. SRI is developed based on two assumptions: 1) it is possible to observe internal processes in much the same way as one can observe the external real-world events, and 2) human can verbalize their internal thought processes. SRI has demonstrated the considerable potential when studying cognitive strategies and learning processes, which suggested that it is an effective way to gain learners' interpretation of events and their thinking at a particular point in time.

The latest research has discovered the employment of biofeedback like eye-tracking data as stimuli for SRI is an available approach to explore learning behavior [22]. SRI can stimulate learners to reflect their viewing behavior and vocalize their thinking processes. Aiming at receiving data explicitly, in the current research, we are proposed to explore users' learning behavior with EDs. The heat map and scan path generated from eye-tracking to participants in the experiment were offered as stimuli to remind them of their processing, and collect retrospective feedback to further explore the learning behavior.

D. Studies on Electronic Dictionary

ED service providers, such as Bing by Google, Youdao by Netease and iciba by Kingsoft, have been diversifying the design and expanding the amount of displayable content for ED, to meet users' constant requirements [23]. Many EDs also provide interactive interfaces, pronunciation, glossary notebook, etc. Given the personalized learning needs and

unique using behavior, the users expressed their requirements to have personalized ED products efficiently.

Previous studies on EDs designing have been conducted from two major perspectives: 1) designing, and 2) content. For example, some researchers put it that different fonts may affect the efficiency and learning the effect of EDs; colored functional label in EDs reduce the time of dictionary lookup and make the search more successful. The layout of the EDs final entry page should be dynamic and customizable; Hierarchical arrangement of lexicographic data, flexible presentation modes and highlighting search results can better navigate users and prevent them from getting lost in searching for words [24][25]; For content, although definition is the basic function, example sentences, especially with diversified contexts are helpful for learners, and only multiple examples presented in different contexts were considerably more helpful than definitions. However to what a number of examples should be diversified remained to be studied, which implied that ED designing and content should be further improved.

The existing studies only take ED users as the passive recipient of information provided by ED and focus on helping users improve their look-up behavior [26]. Users prefer to take the initiative to actively interact with EDs, and their cognitive process might vary due to different design of the EDs and affect their using behavior. Studies focus on the interaction between ED design and using behavior has been scares.

III. Research Method

A. Participants

The participants of this study were proposed to be selected from native Mandarin speakers, who are studying as international students from a university located in the south of the United States. The selected sample should obey the following criteria: firstly, all the participants are required to use EDs in various situations frequently to make sure their comprehension of the ED functions. Secondly, all recruited participants should have full or adjusted to full visual acuity to be qualified for the calibration of the eye-tracker. Finally, participants could be willing to perform immediate stimulated recall interview based on their willingness to share their thinking with their eye-tracking heat map and scanning path figures. Due to the above criteria, a total of 48 international students are recruited.

B. Procedure

The procedure of this experiment, shown in Fig.1 was conducted as three steps. Firstly, the pretest survey contains ten items, includes demographic information (gender, age, year in the college), familiarity to EDs (the most frequently used ED product and the frequency of using), the preferential functions and reasons for using EDs based on their prior experiences. Items besides demographic information were on a 1-5 Likert scale.

The eye-tracking device used to conduct the eye-tracking data in this study named Eye Tribe is produced by a Danish company named Eye Tribe Company. Its sampling rate is 30 Hz and 60 Hz mode. Eye gazing and calculated gaze point was the two major input modality. All the collected eye-tracking data was recorded automatically by the eye-tracker and analyzed by the matched software named Eyeproof, for subsequent use in the SRI data collection. Combined with other input devices including a mouse, keyboard, touch and gestures during the eye-tracking test, the behavior of all participants during the test was videotaped with a video camera. The pages of searched words with three EDs were presented as images on a 19-inch monitor. The distance between the monitor and the chin rest of the eye-tracker was approximately 60cm.

Secondly, the heat map and scanning path were generated from the Software Ogama, which record learners' contemporary reading behavior. Thirdly, the stimulated recall approach was introduced in the retrospection process, that learners used their fixation duration, fixation counts and scanning path as stimulus to retrospect their previous reading behavior and offered their comments to teachers afterwards. The approximate time span for each participant completed all the viewing tasks was about 10 minutes. After all the participants completed the eye-tracking task, they were asked to complete an interview with the interviewer on their EDs viewing experience. The interview lasted 5-10 minutes for each participant. Content analysis was employed to investigate the interview data, during which learners' comments on their contemporary using behavior were categorized into some features used for comparison analysis.

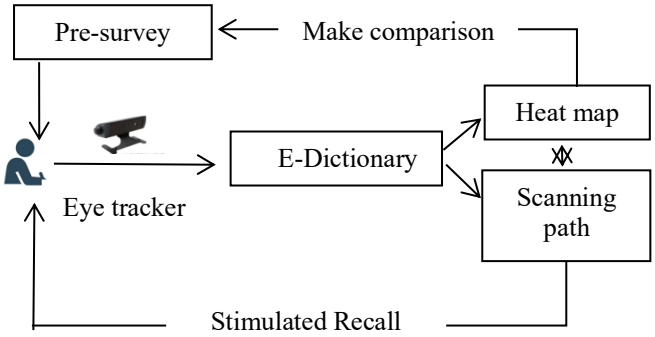

Fig. 1. The procedure of the integration of eye tracking and stimulated recall.

IV. Results

Two images were generated from the eye-tracker: heat map contains fixation count and duration data, and scan path image contains scan pattern. Observed from the heat map, learners' fixation duration was formed mainly on five areas of interest, which are: 1) pictures, 2) interactive links (hyperlinks, voting links); 3) first three example sentences; 4) pronunciation (pronunciation button and Phonetic symbol), and 5) the first and second definition of the target words. Additionally, fixation counts densely landed in four regions: pictures, example sentences, pronunciation (pronunciation button and Phonetic symbol), and definitions.

The scanning path image showed the following observations: the general scan path started from definitions of words, move to pronunciations and derivatives, back to definitions, move to pictures, examples, then move back to definitions. Interestingly, their eyes move back and forth

between the definitions and multimedia sources, which supported the assumption that users tried to make the connection between definitions and multimedia information constantly. Moreover, throughout the scanning path, there are two sets of looking-back-and-forth behavior reflected: the first set is between pictures and example sentences; the second set is between pronunciations and definitions. The first set showed that users were expecting to build connections between pictures and example sentences, and receive better understanding from pictures towards sentences; the second set revealed that users were trying to associate words with their pronunciations to facilitate memorizing by clicking the pronunciation buttons and viewing the phonetic symbols. Specifically, users were clicking through all the available pronunciations (British， American), which indicated that users interested in various versions of pronunciations, hoping to use multiple pronunciations to support pronunciation memorizing, eventually memorize the words.

V. Discussion

According to dual-coding theory, information coming in from both visual and auditory channels contribute to information processing. The attention users paid to multiple multimedia elements showed in two modes on ED pages, physical representation – text (definition, examples sentences, pictures) and sensory representation (visual and auditory materials). When users start viewing the pages, their working memory started to make sense of sounds and images attended to by users. When words and pictures presented to users, the printed words and pictures impinged on the eyes and spoken words impinged on the ears. If the impinged words didn't ring a bell with users, their eyes moved to the pronunciation button to seek for help. Users' attention was drawn to the auditory sensations coming in from the ears when pronunciations played, and they viewed back to words and definitions to better connect the printed words and the way they should be spoken. During the back-and-forth viewing, users' long-term memory started working on integrating relevant prior knowledge to the sense-making process, to assist users to construct verbal and pictorial models, and form deep working memory. Therefore, more multimedia materials in EDs will facilitate information seeking, learning and memorizing for users.

It is also observed that fixation duration formed mainly on first three example sentences, and the first two definitions of the target words. According to Cognitive Load Theory, the amount of information processing demands evoked by the ED pages exceeded the processing capacity of the cognitive system of most users, in another word, they were cognitively overloaded. Cognitive overload is a challenge for ED designing and users. Efficient using of EDs requires both substantial cognitive processing and cognitive-processing-friendly designing for EDs. Therefore restrict maximum 3 example sentences will fit learners' cognitive load according to eye-tracking data.

When viewing the EDs, users are capable of making sense of the presented words and related material and generatively processing the information. Referred to the heat map, there are overlaps among fixation counts, which revealed that when users are viewing the examples, they also viewed the pictures, pronunciations, and definitions. The length of example sentences created a heavy cognitive load, viewing back and forth is the strategy that users tried to obtain support from making connections among multimedia sources, hoping to reduce the cognitive load, and better understand the sentences, which is consisted with the CTML. The five core processes in understanding a word including selecting words, selecting images, organizing words, organizing images, and integrating – these processes place demands on the cognitive capacity of the information-processing system, and they all have to be fulfilled for generative processing.

VI. Implications

The number of example sentences selected into EDs should be taken into consideration. Throughout the study, the findings show a contradiction between survey results and eye-tracking data: the high demand of more example sentences participants advocated in the pre-experiment survey versus the fixation duration mainly marked the first three sentences. Although learners expect more learning contexts to support the understanding of the definition, in fact, their learning capacity is limited. It is observed in our selected EDs that too many example sentences exemplify only a few high-frequency definitions, left some definitions without explaining. In ED design, example sentences should be diversified to cover equivalent contexts and definitions, rather than the amount presented.

The pronunciation provided could be more diversified. Different pronunciations could provide stimulus effect and enhance impinging the targeted words to spoken words and improve pronunciation connection. Typically, EDs provided British and American pronunciation. It will be beneficial if EDs add diversified pronunciations, such as pronunciations from male, female, senior and young contributors for users to choose from. As both visual and auditory channels functioning in information processing, various versions of pronunciation could facilitate different gender and age groups impinging printed words to spoken words, and help with memorizing.

Pictures should be presented closely to the corresponding definitions and example sentences. Based on the contiguity principle in CTML [27], learners learn more effectively with the lighter cognitive load when text and graphics are tightly integrated spatially and temporally, rather than presented separately. In ED designing, it can be embodied as presenting pictures and example sentences on the same screen at the same time. Integrated presentation increases the probability for learn to keep both the linguistic and psychological representations, and build psychological connections. If presented separately, learns to need to consume corresponding cognitive resources to form connections between Semantic and temporal representations.

The pictures selected for EDs need to be representative from two perspectives: 1) universally acceptable in people's routine, and 2) easily connect to definitions and example sentences. Participants claimed that pictures related to their routine contributed more to their learning, and drastically lowered their cognitive load. Despite the multiple images Bing Dictionary provided for the targeted words, they do not coincide with definitions or the contexts of example sentences,

but identical to the first 3-5 images showed in Bing image search for the same word. These pictures rarely match multiple definitions of the target words, much less supporting contexts of multiple example sentences. In CTML, generative processing happens when learners actively integrate prior knowledge into working memory. Selective pictures could facilitate learners building connections between words, definitions, example sentences, and pictures, successfully extract previous knowledge from long-term memory, integrate the processed information with previous knowledge, hence improve learning outcome.

According to eye-tracking and SRI， participants preferred to cooperate the construction and/or modification of the EDs. Many participants left click mark by using the mouse when viewing the ED pages, most of these marks were left on the hyperlinks of extended information, and interactive sections for instant feedback on Youdao dictionary, instances could be seen in the heat map. Users showed strong interest in participating in providing feedbacks like the usefulness of the picture, definition, example sentences provided. EDs, therefore, could be designed as an open system for wide user involvement. Combined with creative thinking skills, EDs as a multimedia learning tool could be not only used as a tool for information searching, but also a collaborative and creative learning tool.

VII. LIMITATIONS

The participant is limited to Chinese Mandarin speakers because the selected EDs are all English-Chinese dictionaries. Besides, only the image of the final entries of EDs was included and studied, while the sound was not explored because of the limitation of eye-tracking. The relationship between two information coding channels was not included in this study. Even though information contains in ED pages entered the information processing system via one channel (visual), but users are capable of transit the received information into the other channel (audio) to process the information. When learners possess enough cognitive resources, information originally presented to one channel could be transported into the other channel. The instance could be seen in the experiment that images showed on the screen was first received by eyes and processed by visual channel, but an experienced user can transfer the received information into sound in mind, and processed in the audio channel.

REFERENCES

[1] Lew, Robert, and Gilles-Maurice De Schryver. "Dictionary users in the digital revolution." International Journal of Lexicography 27, no. 4 (2014): 341-359.

[2] Fageeh, Abdulaziz Ibraheem. "Effects of Using the Online Dictionary for Etymological Analysis on Vocabulary Development in EFL College Students." Theory & Practice in Language Studies 4, no. 5 (2014).

[3] Niitemaa, Marja-Leena, and Päivi Pietilä. "Vocabulary Skills and Online Dictionaries: A Study on EFL Learners' Receptive Vocabulary Knowledge and Success in Searching Electronic Sources for Information." Journal of Language Teaching and Research 9, no. 3 (2018): 453-462.

[4] Lew, Robert, Marcin Grzelak, and Mateusz Leszkowicz. "How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study." Lexikos 23, no. 1 (2013): 228-254.

[5] Baddeley, Alan D. Human memory: Theory and practice. Psychology Press, 1997.

[6] Chandler, Paul, and John Sweller. "Cognitive load theory and the format of instruction." Cognition and instruction 8, no. 4 (1991): 293-332.

[7] Sadoski, Mark. "A dual coding view of vocabulary learning." Reading & Writing Quarterly 21, no. 3 (2005): 221-238.

[8] Liu, Tzu-Chien, and Po-Han Lin. "What comes with technological convenience? Exploring the behaviors and performances of learning with computer-mediated dictionaries." Computers in Human Behavior 27, no. 1 (2011): 373-383.

[9] Mayer, Richard E. "Multimedia instruction." In Handbook of research on educational communications and technology, pp. 385-399. Springer New York, 2014.

[10] Mayer, Richard E. "The promise of multimedia learning: using the same instructional design methods across different media." Learning and instruction 13, no. 2 (2003): 125-139.

[11] Park, Babette, Lisa Knörzer, Jan L. Plass, and Roland Brünken. "Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms." Computers & Education 86 (2015): 30-42.

[12] Majaranta, Päivi, and Andreas Bulling. "Eye tracking and eye-based human–computer interaction." In Advances in physiological computing, pp. 39-65. Springer, London, 2014.

[13] Just, Marcel A., and Patricia A. Carpenter. "A theory of reading: From eye fixations to comprehension." Psychological review 87, no. 4 (1980): 329.

[14] Lai, Meng-Lung, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. "A review of using eye-tracking technology in exploring learning from 2000 to 2012." Educational research review 10 (2013): 90-115.

[15] Crosby, Martha E., Brent Auernheimer, Christoph Aschwanden, and Curtis Ikehara. "Physiological data feedback for application in distance education." In Proceedings of the 2001 workshop on Perceptive user interfaces, pp. 1-5. ACM, 2001.

[16] Lorigo, Lori, Maya Haridasan, Hrönn Brynjarsdóttir, Ling Xia, Thorsten Joachims, Geri Gay, Laura Granka, Fabio Pellacini, and Bing Pan. "Eye tracking and online search: Lessons learned and challenges ahead." Journal of the American Society for Information Science and Technology 59, no. 7 (2008): 1041-1052.

[17] Wang, Qiuzhen, Sa Yang, Manlu Liu, Zike Cao, and Qingguo Ma. "An eye-tracking study of website complexity from cognitive load perspective." Decision support systems 62 (2014): 1-10.

[18] Mason, Lucia, K. Scheiter, and Maria Caterina Tornatora. "Using eye movements to model the sequence of text–picture processing for multimedia comprehension." Journal of Computer Assisted Learning 33, no. 5 (2017): 443-460.

[19] Jamet, Eric. "An eye-tracking study of cueing effects in multimedia learning." Computers in Human Behavior 32 (2014): 47-53.

[20] Rodrigues, Pedro, and Pedro J. Rosa. "Eye-tracking as a research methodology in educational context: a spanning framework." In Eye-Tracking Technology Applications in Educational Research, pp. 1-26. IGI Global, 2017.

[21] Mackey, Alison, and Susan M. Gass. Stimulated recall methodology in applied linguistics and L2 research. Routledge, 2016.

[22] Zhai, Xuesong, Qiansheng Fang, Yan Dong, Zhihui Wei, Jing Yuan, Luca Cacciolatti, and Yalong Yang. "The effects of biofeedback-based stimulated recall on self-regulated online learning: A gender and cognitive taxonomy perspective." Journal of Computer Assisted Learning (2018).

[23] Tono, Yukio. "Application of eye-tracking in EFL learners' dictionary look-up process research." International Journal of Lexicography 24, no. 1 (2011): 124-153.

[24] Lew, Robert, and Patryk Tokarek. "Entry menus in bilingual electronic dictionaries." eLexicography in the 21st century: New challenges, new applications. Louvain-la-Neuve: Cahiers du CENTAL (2010): 193-202.

[25] Yamada, Shigeru. "A test of the proposed framework for reviewing online dictionaries: mw. com, dictionary. com, macmillandictionary. com, dictionary. cambridge. org, and oxforddictictionaries.

com." Dictionaries: Journal of the Dictionary Society of North America 34, no. 1 (2013): 211-224.

[26] Tseng Tseng, Fan-ping. "EFL Students'" Yahoo!" Online Bilingual Dictionary Use Behavior." English Language Teaching 2, no. 3 (2009): 98-108.

[27] Mayer, Richard E. "Multimedia learning." In Psychology of learning and motivation, vol. 41, pp. 85-139. Academic Press, 2002.