Cambridge Open-Review Educational Research e-Journal
Vol. 6, October 2019, pp. 34–49

DOES 'WOW' TRANSLATE TO AN 'A'? EXPLORING THE EFFECTS OF VIRTUAL REALITY ASSISTED MULTIMODAL TEXT ON CHINESE GRADE 8 EFL LEARNERS' READING COMPREHENSION

Yudan Su
University of Cambridge, Cambridge, United Kingdom
ys499@cam.ac.uk

In recent years, the incorporation of multimedia into linguistic input has opened a new horizon in the field of second language acquisition (SLA). In the reading aspect, the advent of virtual reality (VR) technology extends the landscape of reading repertoire by engaging learners with auditory, visual and tactile multimodal input. However, few studies have yet examined the pedagogical potential of VR technology in enhancing learners' reading comprehension. This study aims to fill this gap by assessing the effects of VR-assisted multimodal input on learners' expository reading comprehension. Three classes including 140 Chinese 8th grade EFL students participate in this study, and these classes are randomly assigned into two experimental groups and one control group: VR-assisted multimodal text group, video-assisted multimodal text group and print-based monomodal text group. This study adopts mixed methods methodology and triangulates pre-post-retention tests, questionnaires and interview data to compare three modes of text input on learners' reading performance and explore learners' cognitive processing in the multimodal learning environment. This study is the first attempt to integrate VR technology with input presentation and cognitive processing in second language reading comprehension and offered a new line of theorisation of VR-assisted multimodal learning in the cognitive field of SLA.

Keywords: Multimodal input, Virtual reality, Second language acquisition, Reading comprehension, Cognitive load

Introduction

As the world told becomes the world shown, dynamic and multiple representations by using multimedia tools redefine the conventional print-based text and encourage different conceptualisations and different ways of thinking (Kress and van Leeuwen, 2001). This implies that conventional theories of reading comprehension based on paper texts in second language acquisition (SLA) can no longer give a full account on the way people process multimodal texts including sound, words, images, and movement. It is worth exploring how different modes of input affect readers' understanding as well as how readers process different text input and construct multiple mental models. In this study, I used the term "multimodal text" (Walsh, 2007, p. 26) to encompass a broad concept of engaging in, interacting with, and reflecting upon the text presented by different multimedia. This study is situated in the cognitive account of SLA, aiming to find out the potential impact of multimedia on reading comprehension, specifically how EFL learners decode word meanings and construct mental representations of the screen-based multimodal text in the memory system that differs from the print-based monomodal text. This article is a short summary of an ongoing study that aims to fully gauge the efficacy of multimedia, especially VR technology, in providing multimodal input and enhancing Chinese 8th grade EFL students reading comprehension and offers a new category of multimodal text in the scholarship of SLA.

Theoretical Background

Conceptual Framework

This study draws on the cognitive theory of learning with media (Moreno, 2006) and examines the efficacy of multimodal input in enhancing learners' L2 reading comprehension. The cognitive theory of learning with multimedia adapts from the cognitive theory of multimedia learning (Mayer, 2002), which has been adopted as the theoretical framework by many empirical studies to investigate the effects of multimedia on learners' listening comprehension (Jones & Plass, 2002), vocabulary acquisition (Jones & Plass, 2002; Altarriba & Knickerbocker, 2011) and reading comprehension (Plass, Chun, Mayer, & Leutner, 2003) in the L2 field. The cognitive theory of multimedia learning (Mayer, 2002) incorporates Paivio's dual coding theory, Baddeley's model of working memory and Sweller's theory of cognitive load and the central concept is that the dual input through words and pictures within learners' working memory capacity together with prior knowledge can promote learning outcome (see Figure 1).

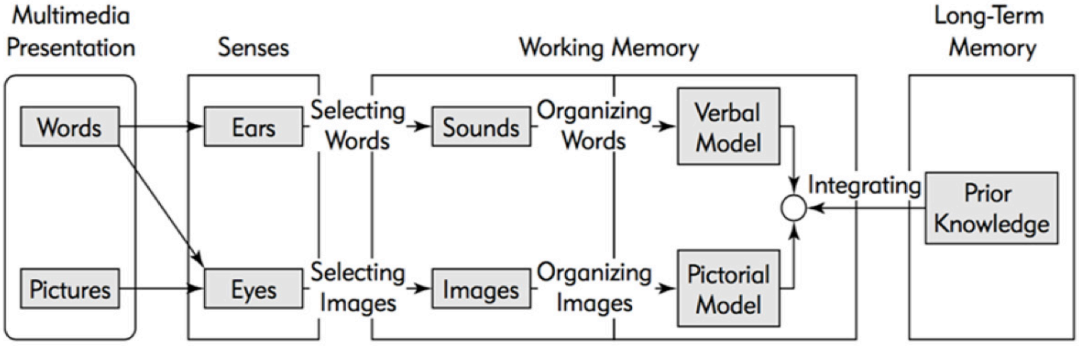

Figure 1. Cognitive theory of multimedia learning (Mayer, 2002)

Moreno (2006) expands Mayer's theory (2002) to include "media such as virtual reality, agent-based, and case-based learning environments" (p. 313) by adding manipulative input that enters into tactile sensory memory (see Figure 2), because the development of multimedia technologies such as VR platform extends multimedia learning to a multidimensional level beyond the scope of 2D verbal and pictorial input. For example, haptic feedback in VR platform allows information to be reinforced through the tactile sensory modality. Hence, VR-assisted input can provide learners with visual presentation, auditory narration and tactile interaction in the reading activity. In brief, this framework shows how multimodal input can promote learners' active cognitive processes through a triple memory model.

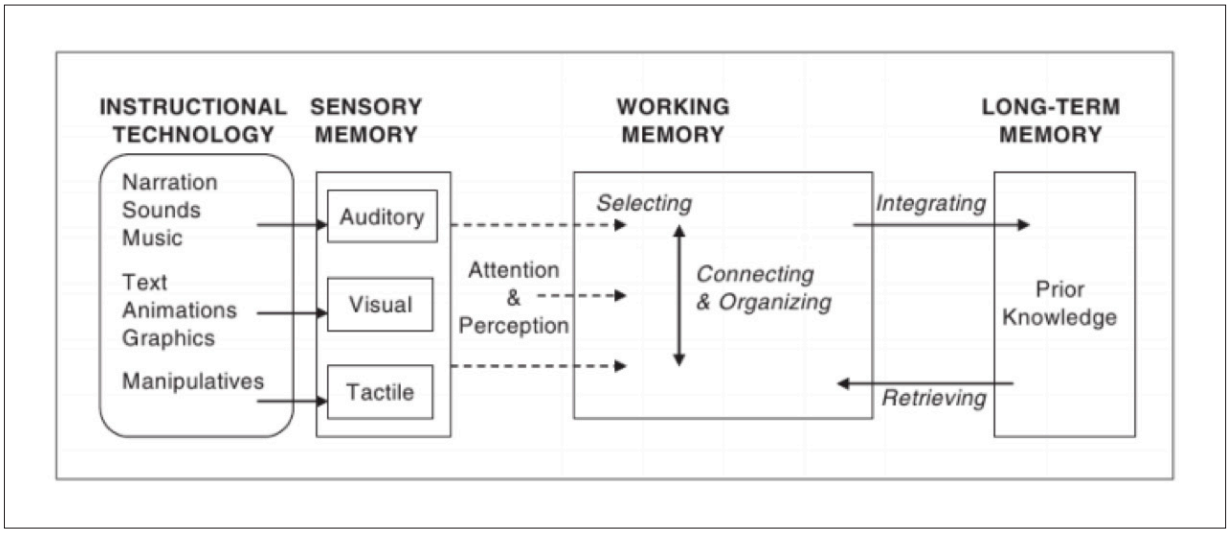

Figure 2. Cognitive theory of learning with media (Moreno, 2006)

Situated in a broader picture of SLA, working memory as a gateway between sensory memory and long-term memory has been found instrumental in facilitating learners' vocabulary acquisition and text comprehension. According to Kintsch and Kintsch (2005), readers' comprehension can be further divided into microstructural and macrostructural levels that serve different functions. The microstructural comprehension means that readers can decode small linguistic units such as word meaning while the macrostructural comprehension requires readers to establish memory representation of the text. Both levels of reading comprehension require learners to identify words meaning, make links between sentences, develop a coherent representation of the content, integrate the textual information with prior knowledge and then organize it in the working memory. Thus, learners' reading comprehension in the multimedia learning environment largely depends on how multimodal input would influence the working memory.

However, working memory has limited capacity to process and store multimodal input and there are mixed findings regarding the effects of multimedia assisted input on learners' cognitive load. On the one hand, multimedia assisted input may exert the modality effect and reduce learners' cognitive load when information is presented in different channels, while on the other hand it may exert the redundancy effect and increase learners' cognitive load when identical information is presented in different channels simultaneously (Mayer, 2005). Since most studies have focused on visual and auditory input, it is of conceptual value to evaluate the effects of visual, auditory and

tactile multimodal input imposed by VR technology on learners' cognitive load and memorisation of textual information. Therefore, the proposed study will examine whether VR-assisted multimodal text could reduce learners' cognitive load and deepen working memory, thereby enhancing reading comprehension in comparison to video-assisted multimodal text and print-based monomodal text.

Literature Review

Multiple studies have applied different modes of multimedia such as pictures, audio, video and subtitles to facilitate learners' L2 acquisition (Lorenz, 2009; Lan and Sie, 2010) by using pre-test and post-test methodology in controlled experimental environment, while the cognition of learners (Neo, 2009; Wiebe & Kabata, 2010) has been less explored to capture a comprehensive snapshot of the effects of multimedia, especially VR technology on the reading aspect of SLA. Thus, this study will bridge the gap by conducting stimulated recall interviews to probe into learners' cognitive processes in the reading activity to understand how multimodal input affect their memorisation of textual information. Furthermore, several studies have explored the effects of auditory and visual input on learners' reading comprehension. Son (2003) found the integration of text with sound and image achieved greater reading comprehension than paper-based text reading. Similarly, Lewandowski, Begeny, and Rogers (2006) found that computer-based reading program with visual and auditory elements improved 3rd grade learners' word recognition, reading speed, and accuracy scores significantly. Segers and Hulstijn-Hendrikse (2008) compared three ways to teach reading comprehension and found that using oral presentation with pictures offered the optimum way to improve learning outcome. It is noted that the majority of previous studies was limited in providing auditory and visual input, and to date there exists a paucity of studies examining the use of VR-assisted multimodal text in the context of L2 reading and no study to date has focused on Chinese EFL beginners' reading comprehension. Therefore, this study will address the research gaps in MCALL and SLA fields by exploring the effects of VR-assisted multimodal input on Chinese 8th grade EFL learners' macrostructural and microstructural reading comprehension from objective performance and subjective cognition.

Research Questions

Situated in the cognitive theory of learning with multimedia, the study attempts to answer the following questions:

1. What are the immediate and delayed effects of input modality (VR-assisted multimodal text, video-assisted multimodal text, print-based monomodal text) on Chinese EFL learners' macrostructural and microstructural reading comprehension?
2. Does input modality (VR-assisted multimodal text, video-assisted multimodal text, print-based monomodal text) affect learners' cognitive load? If so, how does the changed cognitive load influence Chinese EFL learners' reading comprehension?
3. How do Chinese EFL learners perceive the effects of input modality (VR-assisted multimodal text, video-assisted multimodal text, print-based monomodal text) on reading comprehension?

The first question aims to examine the efficacy of multimodal input on learners' reading comprehension at two levels. Screen-based multimodal text including VR-assisted and video-assisted multimodal texts and print-based multimodal text will be compared regarding the effects on macrostructural and microstructural reading comprehension. The second question evaluates whether multimodal input would exert the modality effect or redundancy effect on the working memory, thereby affecting learners' reading comprehension. The third question captures learners' perceptions towards the effectiveness of multimodal input and provides a detailed account of learners' cognitive processes in the multimedia-assisted reading activity. Overall, the three questions aim to fully gauge the efficacy of multimodal text on L2 reading comprehension from objective performance and subjective perception.

Research Design

To address the three research questions, this study takes an important notion from Hitchcock and Hughes (1995) to design the research, starting from ontological assumptions to epistemological assumptions to methodological considerations and instrumentation, which can be briefly summarised in the Table 1.

Table 1
The Integrated Mixed Methods Research Design

	Focus	Data collection	Data analysis
RQ1: the effects of input modality on reading comprehension	Immediate effects	Reading task after each session	SPSS: ANOVA
	Delayed effects	Reading task two weeks after sessions	
RQ2: the effects of input modality on learners' cognitive load	Immediate effects	Cognitive load scale	SPSS: Paired sample t-test
	Delayed effects	Cognitive load scale	
RQ3: learners' perceptions towards the effectiveness of multimodal text	learners' perceptions	Stimulated recall interviews	Coding: descriptive and pattern coding

Research Methodology

This study is situated in the pragmatist paradigm which "is not committed to any one system of philosophy or reality but focuses on the 'what' and 'how' of the research problem" (Creswell, 2003, p.11). The major rationale behind using pragmatism to guide this research is because the focus of this study is to design, implement and evaluate an intervention regarding multimodal text input. Based on the pragmatist paradigm, the study adopts mixed methods methodology to address research questions, which does not only combine separate strengths of relevant quantitative and qualitative methods but also integrate them into a systematic set of data collection methods. The choice of mixed methods methodology derives from the three research questions. The first and second questions compares different input modalities on Chinese EFL learners' reading performance and cognitive load by collecting quantitative data from pre-test, immediate post-test and delayed post-test and scale ratings. The third question probes into participants' perceptions towards multimedia assisted reading experience, thereby encouraging the researcher to utilise a qualitative method such as interview to collect feedback.

The study utilises concurrent triangulation strategy by collecting quantitative data and qualitative data in one phase as shown in Figure 2, although quantitative data will be given more weight over qualitative data in this study. The data will be analysed separately and then synthesised in the interpretation stage. The rationale for adopting this approach is that it can answer the research questions and triangulate quantitative and qualitative results to cross-validate research findings and conclusion.

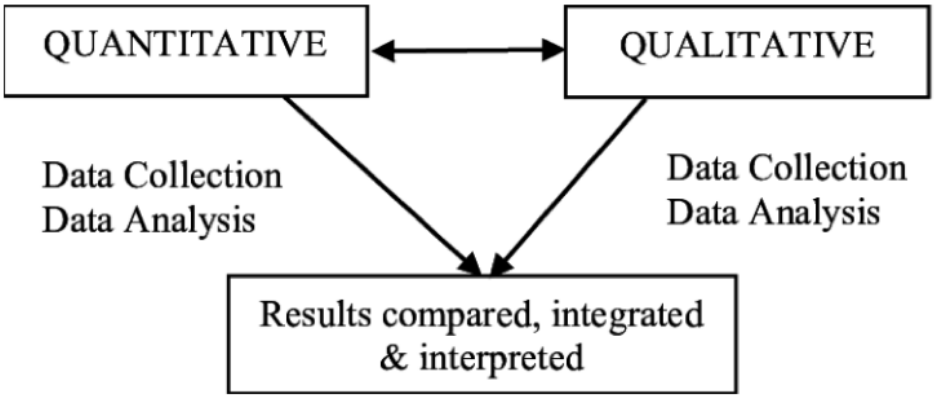

Figure 3. Visual diagram of mixed methods concurrent triangulation strategy (Atif et al., 2013)

Ideally, the controlled experimental research design offers the optimum way to explore the causal relationship between input modality (independent variable) and learners' reading comprehension (dependent variable). However, since the study will take place in a local middle school, it is not feasible to assign subjects into different conditions randomly, which may disturb the normal operation of the target school. Thus, this study adopts a quasi-experimental approach to examine the effects of multimodal input on learners' reading comprehension. While quasi-experimental design is more practical to implement than experimental design, it is more susceptible to threats to internal validity. The target school has streamed classrooms by grouping students based on their academic performance into above average, average or below average. Students within a certain grade range are grouped together as a class. This study addresses the potential threat by selecting three average classes in which students have similar academic performance and language proficiency.

Participants and Sampling

The target sample in this study is Chinese L1-English L2 learners in grade 8 of a middle school. Three classes with a total of 140 students will participate in the study, and the average age of participants is 14 years old. The target school is located in Nanchang, Jiangxi Province where the Ministry of Industry and Information Technology and the provincial government take the lead in building a world-class VR centre. Given the limited time to undertake the research, I collaborate with the school that has been selected as a pilot school to apply VR technology into secondary education and equipped with a VR lab, rather than to find new schools to implement the technology from scratch. This school pays special attention to integrate technology into the curriculum and begins to teach biology, geography and history classes in the VR lab while the VR technology has not been incorporated into English subject yet.

This study adopts non-probability purposive sampling. Due to the restriction of the research site, it is not practical to draw a random sample from all the Chinese 8th grade EFL learners in the target school. In order to address the threats to the validity and generalisability of the research findings posed by non-probability sampling, this study selects three classes from the same average level of academic performance without previous exposure to VR-assisted learning experience that can largely represent the sample. The three classes are randomly assigned into two experimental groups and one control group to control the variables of English proficiency and previous related experience. Experimental group A interacts with VR-assisted multimodal text with visual, auditory and tactile elements, and experimental group B watches video-assisted multimodal text with visual and auditory elements, while the control group C is situated in the traditional reading environment and read print-based text with visual element only. The English teacher involved in this study has more than twelve years of teaching experience with a Master's degree in English and previous experience of using the research apparatus. The teacher only leads the reading activity with minimal involvement in assisting learners' reading.

Research Apparatus and Treatment Materials

Figure 4. Major components of the zSpace platform

zSpace all-in-one computer is used as the apparatus to provide VR-assisted multimodal input for participants in the experimental group A. It mainly consists of three components as shown in Figure 3: a specialised all-in-one computer with a 24-inch HD LCD display, a pair of polarized glasses, and a laser-based interactive stylus. The three components will be activated simultaneously to provide learners with an immersive and interactive learning environment. zSpace has developed modifiable curriculums in 3D models based on textbooks in secondary schools.

zSpace brings visual, auditory and tactile elements together and provides learners with multimodal input. I will use the topic 'the lifecycle of butterfly' to illustrate how zSpace is used in the experiment. The VR-assisted multimodal input consists of visual input that presents 3D animation of different stages in the lifecycle of butterfly and digital text that will pop up on the screen once learners click on the relevant image, auditory input that narrates the digital text on the screen with sound effects and tactile input that enables learners to touch, interact with the butterfly, turn it around 360 degrees and even allow learners to closely observe it as if they were taking it out of the device. For experimental group B, the video-assisted multimodal input consists of visual and auditory elements without haptic feedback. Participants watch videos that illustrate the same content such as the lifecycle of a butterfly with subtitles as the digital text and narration as auditory input in a classroom with a computer and projector. Students in the control group C only receives the visual input of print-based text. Under three conditions, new words in the text are annotated in learners' first language Chinese. The treatment materials are selected from 9th grade lesson plan to ensure that 8th grade learners are unfamiliar with topic knowledge, and learners' language proficiency level and availability of the same content in both VR and video platforms are taken into consideration.

Data Collection

Under the guidance of mixed methods methodology, the study utilises both quantitative and qualitative primary data collection methods to attain methodological diversity and triangulation of findings. Firstly, the study mainly uses reading tasks to assess learners' understanding of the expository text. The reading tasks are formatted in multiple-choice questions and blank-filling questions in order to minimise the effect of subjective judgement that may be caused by open-ended questions and evaluate learners' reading comprehension in an objective way. In each reading task, there are six questions with three on testing microstructural understanding (e.g. the meaning of certain words) and three on macrostructural comprehension (e.g. the main idea of the text). One point is given for a correct answer to each question and the final grade will be objective measure of learners' reading comprehension.

Moreover, the survey instrument consists of three parts. Firstly, participants complete a demographic questionnaire (see Appendix 1) prior to the intervention to get a snapshot of learners' background information including age, gender, English learning experience and previous experience of using multimedia, allowing the researcher to describe the sample accurately. Secondly, a prior knowledge survey (see Appendix 2) is administered to remove those who had already known the domain-specific topic well before the intervention so that the variable of prior knowledge will be controlled. Thirdly, participants will complete a cognitive load questionnaire that focuses on learners' invested mental efforts in the reading task and perceptions of material difficulty. This questionnaire is adapted from the measures of Paas (1992) and Sweller, van Merriënboer, and Paas (1998) and Hwang, Yang, and Wang (2013). This questionnaire consists of eight items in mental effort and mental load dimensions with a five-point Likert rating scale (see Appendix 3). Although the self-reported nature of the rating scale may

appear questionable, it has been applied in multiple studies and demonstrated that people were capable to measure their perceived cognitive load by giving numerical indications (Gimino, 2002; Ayres, 2006).

Lastly, stimulated recall interview as an introspective method will be used to elicit learners' thoughts of cognitive processes involved in the reading activity with a visual or audio prompt (Gass & Mackey, 2013). In this study, six participants in each group will be invited to take part in a stimulated recall interview respectively on a voluntary basis. To ensure the validity of interview questions (see Appendix 4), opinions from the two English teachers have been obtained. A total of three interviews will be recorded for further transcription.

All participants remain anonymous during the data collection process. All data from the three sources mentioned above will be triangulated to compare the effectiveness of VR-assisted, video-assisted and print-based input on students' reading comprehension at macrostructural and microstructural levels and explore participants' perceptions towards multimedia-assisted reading.

Research Procedures

The empirical research project can be largely divided into three stages: pre-intervention, reading intervention and post-intervention. Prior to prevention, participants will be introduced to the study and signed the consent form. Baseline data will be obtained by having students finish a pre-test that includes three expository text reading and the prior knowledge survey. A pilot study has been conducted with nine students in a non-participant class. More annotations of words have been added in the text in alignment with learners' language proficiency and one video used in the experimental group B has been changed to a short one with captions.

In the second stage, the three-week intervention will begin in the target school. The treatment will be offered in three sessions at the rate of one per week. The three sessions will be conducted in student-led group work format, in which three students as a group will be given a reading task to guide the reading process and a teacher will give corrective feedback regarding the reading task before implementing the immediate post-test. During the intervention, participants will read, watch, or interact with three treatment materials. In each session, participants will be required to read, watch or interact with multimodal texts within 20 minutes and finish six questions as the immediate post-test within another 20 minutes. Participants in the experimental group A will engage in the VR-assisted reading task and interact with virtual objects actively while participants in the experimental group B will receive video-assisted input.

In the third stage, participants will fill out the cognitive load questionnaire immediately after finishing the reading tasks. After each session, students will be invited to participate in the stimulated recall interview on the same day, during which they will be encouraged to describe the learning experience and reflect on the usefulness of multimodal text after watching a short clip of video or several video images taken in the session. Two weeks later, a delayed post-test as well as the cognitive load scale will be administered again to evaluate the retention of multimodal text on students' reading performance. Overall, the data collection procedure can be summarised in the Table 2.

Table 2
Timeframe of Data Collection Procedure

Time	Stage of plan	Activities	Data set
Week 1	Pre-intervention	Introduction of the study and research apparatus; informed consent form; demographic questionnaire; prior knowledge survey	Questionnaires
		Pilot study; reading task	Pre-test; fieldnotes
Week 2	Intervention	Session 1; cognitive load questionnaire; stimulated recall interview	Post-test; questionnaire; interview
Week 3		Session 2; cognitive load questionnaire; stimulated recall interview	Post-test; questionnaire; interview
Week 4		Session 3; cognitive load questionnaire; stimulated recall interview	Post-test; questionnaire; interview
Week 6	Post-intervention	Reading task; cognitive load questionnaire	Delayed post-test; questionnaire

Data Analysis

After the fieldwork, quantitative data and qualitative data will be analysed separately and integrated at the interpretation stage of the research. SPSS 24.0 will be used to analyse the quantitative data. A two-way mixed ANOVA will be calculated with time of testing and type of input modality as independent variables and test scores as dependent variable to examine if the three groups' performance is significantly different. If there are significant differences among three groups, one-way ANOVA will be performed to examine the effectiveness of three input modalities on expository text comprehension at each time of testing. Post hoc analyses will be applied to further compare the three groups and find differences at macrostructural and microstructural levels of reading comprehension to examine the causal relationship between input modality and reading performance. Moreover, a paired-samples t-test will be performed to compare learners' cognitive load scale in three conditions and examine whether multimedia assisted input exceeds the working memory capacity and hinders learners' reading or lies within the capacity and improves learners' reading performance.

As for qualitative data, content analysis method will be utilised. Interviews will be transcribed verbatim and coded thematically and analysed by NVivo to answer the third research question regarding learners' perceptions. The interview will be used to capture learners' cognitive processes in the intervention and their evaluation of three input modalities in assisting their expository text reading, which will be helpful to explain the quantitative findings in terms of test scores and cognitive

load scales. The coding will take two steps: (1) descriptive coding that transcribes the recording line by line and provides a detailed inventory; (2) pattern coding that classifies the transcript into focused themes and identifies patterns (Saldana, 2009). At the interpretation stage, the research findings from the two strands will be synthesised and examine the effectiveness of VR-assisted multimodal text in improving EFL learners' reading comprehension from the perspectives of objective learning outcome and subjective individual cognition.

Conclusion

This paper mainly introduces theoretical background, research questions and research design of a study in progress. Upon completion, this research will contribute to the development of MCALL and SLA in three aspects. Firstly, the study incorporates VR technology into L2 reading, extending the scope of multimedia learning in the L2 education field. Secondly, the study offers valuable insight into a new line of theorisation in the aspect of multimodal text, that is screen-based multimodal text and print-based monomodal text, expanding the concept of reading in the digital era. Thirdly, the study presents methodological procedures for implementing VR-assisted reading intervention and sheds light on future pedagogical practice to integrate multimedia technology with language education. Due to the limited time and resources of fieldwork, this study will only expose learners to limited multimedia-assisted input and focuses on expository text reading. Thus, this study will not attempt to make a confirmative conclusion regarding the effectiveness of multimedia technology in L2 education since it takes deliberate, thoughtful, and specific applications for multimedia to be effective and it largely depends on the research design and pedagogical practice. More longitudinal research and careful experimental design need to be done on multimedia-assisted learning within the L2 education field.

References

Altarriba, J., Knickerbocker, H., Trofimovich, P., & McDonough, K. (2011). Acquiring second language vocabulary through the use of images and words. In *Applying priming methods to L2 learning, teaching and research*, 21-48.

Atif, A., Richards, D., & Bilgin, A. (2013). A student retention model: empirical, theoretical and pragmatic considerations. In *24th Australasian Conference on Information Systems (ACIS)* (pp. 1-11). RMIT University.

Berg, B. L. (2007). *Qualitative research methods for the social sciences*. London: Pearson.

Brinton, L. J. (1991). The origin and development of quasimodal 'have to' in English. In *Workshop on the Origin and Development of Verbal Periphrases*, 10th International Conference on Historical Linguistics (ICHL 10), Amsterdam.

Chow, A., Andrews, S., & Trueman, R. (2007). A 'Second Life': can this online, virtual reality world be used to increase the overall quality of learning and instruction in graduate distance learning programs. In *Proceedings of the Association for Educational Communications and Technology International Convention* (Vol. 2, pp. 75-83). Association for Educational Communications and Technology, Bloomington, IN.

Clarke, D., Keitel, C., & Shimizu, Y. (2006). The learner's perspective study. In *Mathematics Classrooms in*

Twelve Countries:The Insider's Perspective, 1, pp. 1 - 14.

Clark, R. E. (2001). *Learning from media.* Greenwich, CT: Information Age Publishing.

Cohen, L., Manion, L., Morrison, K., & Morrison, R. B. (2007). *Research methods in education.* London: Routledge.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In *Handbook of mixed methods in social and behavioral research*, 209, 240.

DeHaan, J., Reed, W. M., & Kuwanda, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. In *Language Learning & Technology, 14*(2), 74-94.

Duffy, J. (1985). *Effects of instruction in noting and using an author's pattern of writing on sixth graders' understanding and memory of expository text* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.

Freina, L., & Ott, M. (2015). A Literature Review on Immersive Virtual Reality in Education: State of the Art and Perspectives. In *eLearning & Software for Education*, (1).

Gottlieb, M. (1999). *The language proficiency handbook: A practitioner's guide to instructional assessment.* Springfield, IL: Illinois State Board of Education.

Hitchcock, G., & Hughes, D. (1995). *Research and the teacher: A qualitative introduction to school-based research.* Psychology Press.

Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. In *Computers & Education, 69*, 121-130.

Jamaludin, A., Ho, C. M. L., & Chee, Y. S. (2007). Argument-Based Negotiation and Conflict Resolution through Enactive Role Play in Second Life. In *Frontiers in Artificial Intelligence and Applications*, 162, 561.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. In *Educational researcher, 33*(7), 14-26.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. In *The modern language journal, 86*(4), 546-561.

Lan, Y. F., & Sie, Y. S. (2010). Using RSS to support mobile learning based on media richness theory. In *Computers & Education, 55*(2), 723-732.

Lorenz, B., Green, T., & Brown, A. (2009). Using multimedia graphic organizer software in the prewriting activities of primary school students: What are the benefits?. In *Computers in the Schools, 26*(2), 115-129.

Mackey, A., & Gass, S. M. (Eds.). (2011). *Research methods in second language acquisition: A practical guide.* John Wiley & Sons.

Mackey, A., & Gass, S. M. (2013). *Stimulated recall methodology in second language research.* Routledge.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85-139). Academic Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. In *Educational psychologist, 38*(1), 43-52.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning, 43*.

McCormick, S., & Zutell, J. (2015). *Instructing students who have literacy problems* (7th ed.). Boston, MA: Allyn & Bacon.

Moreno, R. (2006). Learning in high-tech and multimedia environments. In *Current directions in psychological*

science, 15(2), 63-67.

Morgan, D. L. (2014). Pragmatism as a paradigm for social research. In *Qualitative Inquiry, 20*(8), 1045-1053.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. In *Journal of educational psychology, 84*(4), 429.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. In *Computers in Human Behavior, 19*(2), 221-243.

Roehling, J. V., Hebert, M., Nelson, J. R., & Bohaty, J. J. (2017). Text structure strategies for improving expository reading comprehension. In *The Reading Teacher, 71*(1), 71-82.

Saldaña, J. (2015). *The coding manual for qualitative researchers.* Sage.

Segers, E., Verhoeven, L., & Hulstijn-Hendrikse, N. (2008). Cognitive processes in children's multimedia text learning. In *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 22*(3), 375-387.

Son, J. B. (2003). A hypertext approach to foreign language reading. Australian Review of Applied Linguistics. In *Supplement Series, 17*(1), 91-110.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. In Educational psychology review, 10(3), 251-296.

Walsh, M., Asha, J., & Sprainger, N. (2007). Reading digital texts. In *The Australian Journal of Language and Literacy*, *30*(1), 40.

Appendix A: Demographic questionnaire

1. What is your gender?

- ○ Female
- ○ Male
- ○ I prefer not to say

2. What is your age in years?

3. How long have you studied English at school?

4. Which aspect of English learning do you find most difficult?

- ○ Listening

- ○ Reading
- ○ Writing
- ○ Speaking

5. Which type of reading have you learned at school?

- ☐ Narrative reading 记叙文
- ☐ Descriptive reading 描写文
- ☐ Expository reading 说明文
- ☐ Persuasive reading 议论文

6. How difficult is it for you to read something for academic purposes (such as school homework/exam)

- ○ Very difficult
- ○ difficult
- ○ not very difficult
- ○ not at all difficult

7. What kind of activity will you do before reading?

- ○ Search background information online
- ○ Watch videos
- ○ Brainstorm
- ○ Check vocabulary
- ○ Other, please specify

- ○ Yes
- ○ No
- ○ I don't know

9. Have you used virtual reality (VR) technology to learn English or other subjects?

- ○ Yes
- ○ No

Appendix B: Prior Knowledge Survey

Please answer the following questions about the lifecycle of butterfly.

1. Please explain what is lifecycle.
2. Please list the four stages of a butterfly's lifecycle.
3. Please write down the Chinese translation of caterpillar: 

4. Please write down the Chinese translation of chrysalis: 

5. Please write down what will happen after the adult butterfly stage.

Please rate your knowledge of the lifecycle of butterfly and circle the number of the item that applies to you.

1. Very little
2. Between very little and average(很少和中等之间)
3. Average
4. Between average and very much(中等和很多之间)
5. Very much

Appendix C: Cognitive load questionnaire

Please rate the following statement from 1 (totally disagree) to 5 (totally agree).

Mental load

1. The learning content in this learning activity was complicated.
2. The instructional way in the learning activity was difficult to follow and understand.
3. It was troublesome to answer the questions in this learning activity.
4. During the learning activity, the way of content presentation was hard to follow.

Mental effort

1. I felt frustrated answering the questions in this learning activity.
2. I did not have enough time to answer the questions in this learning activity.
3. I need to put lots of effort into answering the questions in this learning activity.
4. I have tried my best to answer the questions in this learning activity.

Appendix D: Interview questions

Questions	Rationale
What aspects of the VR/video assisted reading activity are most memorable to you?	Encouraging learners to reflect on the reading process
How would you describe your reading experience with VR/video?	
Does the VR/video assisted reading activity help you know more about the reading topic? If so, how does it help you read expository text?	Asking learners to examine the effectiveness of multimodal text reading
Do you think you would use VR/video again to learn English? If so, in what aspects do you plan to use it (vocabulary, writing, listening, storytelling, etc.)?	
Compared with using paper and pen to read print-based text, what do you think are the advantages of using VR/video in the reading activity?	Asking learners to identify advantages and disadvantages of multimedia in reading by comparing traditional print-based text reading and multimodal text reading
Compared with using paper and pen to read print-based text, what do you think are the disadvantages of using VR/video in the reading activity?	