**Study plan**

**2.1 Project Background**

**Vocabulary inferencing strategy in reading**

Learners need to obtain thousands of new word forms and meanings in order to understand a foreign language (Nation, 2006). They still frequently meet unfamiliar words even after years of study during reading. To cope with this, studies suggest that readers adopt various methods to figure out the meanings of the words. These methods include consulting dictionaries, noting them down to further consult with teachers, inferring meanings from context or even ignoring them (Fraser, 1999; Harley & Hart, 2000; Sanaoui, 1995). Lexical inferencing strategy has been reported as the most popular one used by L2 learners among different word-learning strategies (Nassaji, 2003) It is defined as “involves making informed guesses as to the meaning of an utterance in light of all available linguistic cues in combination with the learner’s general knowledge of the world, her awareness of context and her relevant linguistic knowledge” (Haastrup, 1991).

It is quite crucial for L2 reading comprehension (Nassaji, 2002, 2003). Apart from this, this strategy also leads to the immediate comprehension in speaking or listening context (Paribakht, 2004; Qian, 2004). L2 learners who obtain this skill can perform better and understand text easier than unskilled learners, owing to their higher level of syntactic and semantic processes (Nassaji, 2003).

However, effective inferencing during reading is not easy (Bensoussan & Laufer, 1984; Kelly, 1990; Prince, 1996; Schatz & Baldwin, 1986; Shu, Anderson, & Zhang, 1995). L2 learners still meet obstacles if they want to successfully infer the meaning of new words from contexts (Nassaji, 2003). The same situation is found in our EAP classroom and students heavily rely on translation software. This is not good for both timed exam and developing their independent learning. The success of inference highly relies on the matureness of the strategy used and this implies the necessity for learners to use, evaluate and develop the strategy they used in reading (Chamot, 2004; Fukkink & de Glopper, 1998; Kuhn & Stahl, 1998; Walters, 2004, 2006). Comparing with passive encoding, like dictionary consultation, active generation of a target word should better improve memory based on the study on the generation effect (Bertsch et al., 2007; Mulligan, 2004).

**Dynamic assessment**

DA can be used for vocabulary learning and diagnosis. It provides diagnostic information about learner abilities, and at the same time, promotes learner development as a mediated process of individual transition from other-regulation toward self-regulation (Lantolf & Thorne, 2006). DA is based on the Zone of Proximal Development theory (ZPD), defined as "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). Some research has applied DA to classrooms and explored its effect on L2 development (Davin, 2013, 2016; Herazo et al., 2019; Lantolf & Poehner, 2011; Poehner & Leontjev, 2020). In most classroom-based DA literature, given that teachers acting as mediators should take care of multiple learners simultaneously, it was pre-scripted and graduated mediation that was used to make it possible for mediators to conduct a systematic diagnosis of learning and promote development in large-sized classes.

Meanwhile, there is a body of literature that has utilized DA for L2 vocabulary learning. These studies have indicated that DA can be used as an effective means for vocabulary learning (Andujar, 2020; van der Veen, Dobber, & van Oers, 2016). In some studies, DA was used to assist learners in gaining information about unknown words while reading a text (Ebadi et al., 2018; Rassaei, 2020).

**Artificial intelligence chatbots for dynamic assessment**

Encouraging learners to adopt a positive attitude towards vocabulary learning can be of importance. Learners’ attitude has always been a concerning issue in vocabulary assessment (Nation, 2007). Learners’ positive attitude towards vocabulary learning via computer technology has been well-documented in the literature (Esit, 2011).

With the assistance of an AI agent that is able to respond to a learner’s errors in an automated manner, scholars in the field of language assessment have attempted to take advantage of AI technology when implementing assessment (e.g., Ai, 2017; Heift, 2017; Shute & Ventura, 2013). Specifically, open-source chatbot builders such as XIPU AI, which is described as “XIPU AI (君谋) Platform developed by Learning Mall is built on the OpenAI GPT model. With its array of innovative features, users are empowered to engage with an AI assistant in an authentic and interactive manner.” (XJTLU, 2023) and was launched on 25th September in 2023. With the help of AI chatbots, the process of DA can be implemented with every students in the classroom. In addition, XIPU AI offers an automatic transcript function that can be used as evidence of learner performance.

**2.2 Aims and Objectives**

**Study Aim**

This quasi-experimental study aims to evaluate the effectiveness of Chatbot-Assisted Dynamic Assessment (CA-DA) for vocabulary learning among intermediate-level Chinese EFL learners at a transnational university. Specifically, the study seeks to determine how CA-DA can enhance learners' ability to infer the meanings of unfamiliar words encountered during reading.

**Study Objectives**

1. To compare the effectiveness of the CA-DA method and dictionary consultation (DC) strategy for vocabulary acquisition and retention.
2. To explore how CA-DA, based on dynamic assessment principles, can support learners in developing lexical inferencing skills by providing mediational assistance, such as using co-text, discourse, syntactic knowledge, and morphology.
3. To assess the potential of using XIPU AI as an automated dynamic assessment tool, providing cost-effective, one-on-one interaction for learners practicing lexical inferencing skills.

**2.3 Research Question(s) and Methodology**

Q1: What extent do students improve their vocabulary learning (receptive and productive) through CA-DA?

Q2: How can CA-DA be applied to diagnose vocabulary learning?

Q3: What are students’ perception and attitudes toward the role of CA-DA in vocabulary development?

Please fill in this part with information as below:

a. Specific indications of what will be done, when it will be done, and who it will be done to (how will participants be selected)

**Participants**

Due to the focus on intermediate L2 English learners in this study, we employ judgmental sampling for participant selection. The Vocabulary Size Test (VST) will be administered to 12 groups of students enrolled in intermediate-level courses at XJTLU, based on their placement test results. These groups were evenly divided, with 6 groups assigned to the experimental group (taught by Lin and Anning) and 6 groups to the control group (taught by Feiran and Jiahang). The experimental group receives CA-DA treatment, while the control group undergo DC treatment. To ensure fairness, all students in each class complete the same tasks; however, only the results from the designated participants will be analysed.

**Instruments**

***Pretest of target words:*** We will select 6 target words from the teaching syllabus. Before the treatment session, we pre-test our participants’ knowledge of target words and to ensure they do not know these words. They will be asked to indicate whether they know the meaning of the words or not. When they checked yes to an item, for verification they have to write the meaning of that word in Chinese/English.

***Reading texts selection:*** For smooth lexical inferencing, no more than 2% of the text’s words should be unfamiliar to the learners (Laufer, 2013). To identify the texts which satisfied this precondition, we use and adapt the reading materials from the teaching materials designed for intermediate students by the School of Languages in XJTLU, including the 4 target words that are manipulated to appear at least once in each text. Furthermore, the texts are analysed by Vocabulary Profiler (Smith, 2024) using British National Corpus (BNC) lists, which were built from the text files for Paul Nation's Range program.

A panel of four language lecturers who teach intermediate courses subjectively judged the texts in terms of interest for the participants. When three lecturers agreed on the appropriateness of a text, it will be submitted to judgment of difficulty level.

***Experimental sessions:*** During the two experimental sessions implemented using learners’ normal classroom time in one week, participants are asked to read texts and to identify the meaning of underlined target words. The CA-DA group receives graduated chatbot assistance, while the control group utilizes dictionary only. The two groups are allotted the same amount of time, around 25minutes, for each treatment session.

***Posttest and interactions transcribed*:** A posttest is implemented immediately after the second experimental session on the same day, and a delayed posttest is conducted two weeks after the second session to examine vocabulary retention. The results of the tests are statistically analyzed for each group. Additionally, interactions between learners and the chatbots for the CA-DA sessions are transcribed verbatim and both quantitatively and qualitatively analysed to reveal implications for vocabulary learning and diagnosis of learner abilities.

***Interview*:** Semi-Structured Interview/ Focus Group Interview will be conducted based on the final list of students who are willing to take part in. The process will be recorded and the recording will be used for transcript analysis.

b. Whether permissions from other parties (e.g., employers, owners of venues, the police, etc) have been gained

The research does not require any permission from other parties. The research participants are students at XJTLU. The consent from line managers has been gained.

c. Whether and how people will be informed of the study and how consent will be gained (whether the form will be signed, or it will be agreed to as part of proceeding with an online survey)

The participant information sheet will be explained, and we will leave around 10 minutes in class to explain the research and ask for their willingness to take part in it. The consent form will be signed.

d. Where the data will be held and when and for how long and what security arrangements have been made.

All the data will be stored in an online folder at the lead researcher's OneDrive account at XJTLU, and only the research team (Lin Ma, Anning Yang, Feiran Xu and Jiahang Li) will be able to access it. All the files containing the data will be password-protected. Participants’ data held by the research team will be destroyed 5 years following the completion of the study by the lead researcher. Files that are not required for long-term preservation will be disposed of securely at the end of the project.

e. Whether personal information will be collected or not (with a note that if none is to be collected then the consent form should be adjusted to reflect the fact that it will be impossible to remove participants data once they have submitted it)

All tests which may be conducted via LMC or Wenjuanxing, interviews, the identities of the participants will be known and accessed by the research team.

f. If personal information is collected what steps will be included to ensure that any de-identification is successful

All the collected data will be treated with strict confidentiality. In order to better trace the data, for all tests, interviews, the identities of the participants will be known and accessed by the research team for this study only. All the files containing personal information data will be password protected.

**N.B.** Please use the numbering (2.3.a, 2.3.b, and so on) when writing this section of the study plan. This will make it easier for the reviewers to check if you have addressed each of these issues. Also, please make sure that you address each and every point in the above list.

**2.4 Work Plan and Timeline**

Please provide sufficient detail in this section.

***Week 0-1***

***Pretest of target words:*** We will select 6 target words from the teaching syllabus. Before the treatment session, we pre-test our participants’ knowledge of target words and to ensure they do not know these words. They will be asked to indicate whether they know the meaning of the words or not. When they checked yes to an item, for verification they have to write the meaning of that word in Chinese/English.

***Week 0***

***Reading texts selection:*** For smooth lexical inferencing, no more than 2% of the text’s words should be unfamiliar to the learners (Laufer, 2013). To identify the texts which satisfied this precondition, we use and adapt the reading materials from the teaching materials designed for intermediate students by the School of Languages in XJTLU, including the 4 target words that are manipulated to appear at least once in each text. Furthermore, the texts are analysed by Vocabulary Profiler (Smith, 2024) using British National Corpus (BNC) lists, which were built from the text files for Paul Nation's Range program.

A panel of four language lecturers who teach intermediate courses subjectively judged the texts in terms of interest for the participants. When three lecturers agreed on the appropriateness of a text, it will be submitted to judgment of difficulty level.

***Week 1-2***

***Participants selection***

Students will be introduced the project and sign the consent form based on their willingness. Because this study focuses on intermediate L2 learners of English, we use judgemental sampling for participant selection. The Vocabulary Size Test (VST) is administered to 12 groups of students who were enrolled in intermediate level course via placement test in XJTLU. Study participants groups are randomly assigned to an experiment group (6 groups of original teaching classes) and a control group (another 6 groups of original teaching classes). The experiment group receive CA-DA treatment and the control group receive DC treatment. In order to keep equality, all students in the same class will perform the same task and only the participants’ results will be analysed.

***Week 3-4***

***Experimental sessions:*** During the two experimental sessions implemented using learners’ normal classroom time in one week, participants are asked to read texts and to identify the meaning of underlined target words. The CA-DA group receives graduated chatbot assistance, while the control group utilizes dictionary only. The two groups are allotted the same amount of time, around 25minutes, for each treatment session.

***Week 3-6***

***Posttest and interactions transcribed*:** A posttest is implemented immediately after the second experimental session on the same day, and a delayed posttest is conducted two weeks after the second session to examine vocabulary retention. The results of the tests are statistically analyzed for each group. Additionally, interactions between learners and the chatbots for the CA-DA sessions are transcribed verbatim and both quantitatively and qualitatively analysed to reveal implications for vocabulary learning and diagnosis of learner abilities.

***Week 6***

***Interview*:** Semi-Structured Interview/ Focus Group Interview will be conducted based on the final list of students who are willing to take part in.