System

Do AI chatbots impact motivation? Insights from a longitudinal study

--Manuscript Draft--

Manuscript Number:	
Article Type:	VSI: AI in language education
Section/Category:	
Keywords:	Motivation; AI; ChatGPT; Chabot; SRL; Piecewise growth curve models
Corresponding Author:	Meng Liu Beijing Foreign Studies University Cambridge, Cambridgeshire UNITED KINGDOM
First Author:	Meng Liu
Order of Authors:	Meng Liu
	Hayo Reinders
Abstract:	This study investigates whether chatbots designed to engage in conversations about self-regulation of learning impact motivation. Twenty-four students in an academic listening course at a university in China were asked to engage with interactive learning journal chatbots over a period of 15 weeks. In weeks 1–8 they engaged in pre-scripted chatbot conversations and in weeks 9–15 in conversations with an AI-powered chatbot. Piecewise mixed effects models were used to capture changes in learners' motivational trajectories before and after the introduction of generative AI. Our findings suggest the participants experienced a significant upward trend in motivation after the introduction of generative AI. These results indicate that 1) generative AI may positively impact motivation, and 2) more so than pre-scripted chatbots.
Suggested Reviewers:	Ali Al-Hoorie Royal Commission for Jubail and Yanbu hoorie.ali@gmail.com Dr Ali Al-Hoorie is a leading expert on motivation and technology-assisted language learning.
	Yongyan Zheng Fudan University yongyanzheng@fudan.edu.cn Prof Yongyan Zheng is a leading expert on motivation who is also familiar with the Chinese context.
	Phil Hiver Florida State University phiver@fsu.edu Dr Phil Hiver is a leading expert on language learning motivation.
	Abdullah Alamer King Faisal University alamer.aaa@gmail.com Dr Abdullah Alamer is a leading expert in quantitative research, motivation and computer assisted language learning.

Title Page (with Author Details)

Do AI chatbots impact motivation? Insights from a longitudinal study

Meng Liu*, Assistant Professor, Beijing Foreign Studies University, China
mengliu@bfsu.edu.cn

Hayo Reinders, Professor, Anaheim University, USA

*Corresponding author

Do AI chatbots impact motivation? Insights from a longitudinal study

Abstract

This study investigates whether chatbots designed to engage in conversations about self-regulation of learning impact motivation. Twenty-four students in an academic listening course at a university in China were asked to engage with interactive learning journal chatbots over a period of 15 weeks. In weeks 1–8 they engaged in pre-scripted chatbot conversations and in weeks 9–15 in conversations with an AI-powered chatbot. Piecewise mixed effects models were used to capture changes in learners' motivational trajectories before and after the introduction of generative AI. Our findings suggest the participants experienced a significant upward trend in motivation after the introduction of generative AI. These results indicate that 1) generative AI may positively impact motivation, and 2) more so than pre-scripted chatbots.

Keywords: Motivation; AI; ChatGPT; Chabot; SRL; Piecewise growth curve models

1 Introduction

What role will Artificial Intelligence (AI) play in education? And what role should it play? Ignoring the considerable current hype around the topic, relatively little is known about its affordances (its potential benefits and drawbacks; Reinders & Hubbard, 2013), how these align with people's pedagogical aspirations, or how they can be integrated into a particular learning context. With a few exceptions current research is limited largely to user perceptions of short-term interventions that are mostly not systematically designed and not theoretically driven (Jeon, 2024). What is needed instead is a careful investigation of a particular area where AI can be hypothesized to offer an advantage over current practices. One such area is in supporting learners' ability to self-regulate their learning, both inside and outside of the classroom. AI chatbots in particular offer learners the opportunity to engage in personalized conversations about their learning goals, reflect on their experiences, and plan for subsequent learning, among others (e.g., Hew et al., 2023). This is not dissimilar from the type of individualized interaction that would be had with a language advisor or language counsellor (Kato & Mynard, 2015), with such added benefits as lower costs, anytime-anywhere access, a potentially lower affective barrier (e.g., by not feeling embarrassed to share one's shortcomings), and the ability to analyze the results in order to personalize subsequent learning. As will be discussed in the following sections, a focus on self-regulated learning skills has a number of well-researched benefits, not least of which is the impact on learner motivation, which in turn has been widely shown to impact language learning outcomes. A reasonable question to ask then, is whether engaging with a chatbot about one's learning can impact motivation. A related question is what type of interaction is most beneficial. Is a pre-scripted approach better (i.e., one whereby questions are design and sequenced by a teacher or researcher and offered through a traditional chatbot) or is an open conversation better (such as facilitated by generative AI)?

2 Literature Review

2.1 Theories of Language Learning Motivation

Motivation has long been recognized as an influential factor in learners' engagement, persistence, and success in language learning (Dörnyei et al., 2015; Dörnyei & Ushioda, 2021). Several theories have been proposed to conceptualize and explain language learning motivation, such as the L2 motivational self system theory (Dörnyei, 2005, 2009), which is the most widely applied domain-specific theory of motivation (Liu, 2024). An important insight from this literature is the dynamism and context-dependency of motivation, highlighting its sensitivity to the learning environment and individual learner differences (e.g., Dörnyei et al., 2015; Reinders et al., 2022).

Another prominent theory of motivation is self-determination theory (SDT; Ryan & Deci, 2017), a theory from psychology influential in both language learning research and beyond (Al-Hoorie et al., 2022). Basic psychological needs theory (Ryan & Deci, 2017), one of the mini-theories of SDT, posits that human beings have a basic need for autonomy (i.e., self-directed control and choice), relatedness (i.e., connection and belonging), and competence (i.e., feeling capable and effective). When these psychological needs are met, intrinsic motivation can be generated and enhanced, whereas when these needs are thwarted, intrinsic motivation is dampened (Vansteenkiste et al., 2020). Empirical research in the field has shown that supporting learners' basic psychological needs can foster their intrinsic motivation in language learning (e.g., Joe et al., 2017). It is worth noting that contrary to common assumptions, providing structure in the learning environment can actually enhance autonomy and motivation. For example, Oga-Baldwin and Nakata (2015) investigated the role of structure in enhancing autonomy and motivation among Japanese students in foreign language classes. Their findings suggest that a learning environment with a well-defined structure can nurture students' autonomy and motivation.

2.2 Self-regulated Learning

Self-regulated learning (SRL) refers to learners' ability to control, monitor, and regulate their own learning process independently to achieve learning goals (Pintrich, 2000; Zimmerman, 2008). In one of the most widely applied models, Zimmerman (2000) divides SRL into three phases: forethought, performance, and self-reflection. The forethought phase is where learners prepare for learning, engaging in activities such as goal setting and strategic planning (Cleary & Zimmerman, 2004). The performance phase involves using strategies such as self-monitoring to keep track of one's learning progress. The self-reflection phase refers to when learners review and reflect on their learning experience, which can in turn impact subsequent forethought processes (Zimmerman, 2013). Although Zimmerman's model (2000) provides a solid foundation for understanding SRL, alternative models have also been proposed to offer complementary perspectives. For instance, Boekaerts et al. (1999) highlight the emotional dimension of SRL whereas Pintrich (1995) underscores the influence of environmental factors on SRL strategies.

In the context of language learning, interest in learning strategies has a long history (see Rose et al., 2018 for a systematic review). Within this field of research, intervention studies embedded with SRL principles (i.e., strategy instruction research) continue to be a flourishing area. According to Ardasheva et al. (2017)'s meta-analysis, strategy instruction has large effects for both language learning and SRL. Notably, for SRL, technology-delivered strategy instruction can be equally effective or even more so than teacher-delivered strategy instruction (Ardasheva et al., 2017).

2.3 Chatbots for Language Learning and SRL

Chatbots have been recognized as a useful learning aid to facilitate various domains of language learning, ranging from speaking (e.g., Ayedoun et al., 2019) to grammar (e.g., Kim et al., 2019). The pedagogical focus of chatbots also varies – some studies focus on the role of chatbots as conversational partners (e.g., Kim, 2016) while others use them to answer students' questions about the subject (e.g., Wang et al., 2017). Chatbots have been found to increase interaction frequency (Goda et al., 2014) and improve communication strategies (Kim, 2016). In terms of social affordances, chatbots have been found to encourage self-disclosure (Goda et al., 2014) and personal experience sharing (Xu & Warschauer, 2020).

Compared with the focus on language skills development, chatbots' potential in fostering learners' ability to regular their own learning has received less attention. Only a few studies have explored the use of chatbots for this purpose and within the few that did, the focus was primarily on goal setting. Du et al. (2021) explored how a rule-based chatbot helped students set learning goals before starting their online course during Covid-19. Students were prompted by the chatbot to consider their learning objectives and expectations. Learners benefited from the chatbot in gaining greater clarify of their goals and becoming more aware of the benefits of goal setting. In another study using rule-based chatbots, Hew et al. (2023) implemented a goal-setting chatbot in an online course and a learning buddy chatbot in in a EFL listening course at a public university in Asia. Learners reported generally positive experience with both chatbots, perceiving them as easy to use and useful.

In the meantime, how chatbots impact students' motivation to learn a language remains an underexplored area (Bibauw et al., 2019). The few studies in this strand produced mixed findings. Implementing chatbots as language practice tools, Fryer et al. (2017) evaluated the extent to which students' interest in language learning was stimulated and sustained by chatbots and found a novelty effect – learners' interest dropped after the early stages in the course. In contrast, Lee et al. (2011) designed a course where intelligent robots engaged in role play conversations with learners and found that interacting with the robot increased students' satisfaction, interest, confidence, as well as language learning motivation. In a more recent study, Jeon (2024) deployed customized chatbots in an English as foreign language course and identified a mixture of opportunities and constraints afforded by the chatbots. Specifically, students' motivation was found to be influenced by a variety of factors – the learners' perceived linguistic competence, the chatbot's pedagogical value and technological affordances, and learners' perception of chatbots as authentic speakers. As a result, whether these factors facilitated or hindered the learner's engagement with the chatbot was found to be highly individualistic.

Across this line of research, the technological limitations of traditional chatbots emerged as a common theme. Specifically, traditional chatbots have been criticized for, among others, unnatural computer-generated voices (Goda et al., 2014), nonsense outputs (Fryer et al., 2019), lack of affective or visual cues (Gallacher et al., 2018), and limited capacity to engage in spontaneous conversations (Hew et al., 2023), all which may dampen learners' interest and engagement in learning.

The emergence of large language models such as GPT (generative pre-trained transformer) open up new opportunities for both research and practice in language learning and teaching (Chen et al., 2020; Han, 2024; Kohnke et al., 2023; Zou et al., 2023). As demonstrated in recent literature (Kohnke et al., 2023), ChatGPT, one of OpenAI's most advanced large language models, has an impressive ability to realistically mimic human conversation, which potentially addresses the technological limitations of that traditional chatbots have been criticized for .

Taken together, our literature review shows that while the existing literature provides valuable insights into the use of chatbots for language learning and SRL, several gaps can be observed that warrant further attention. First, although a limited number of studies have explored chatbots' impact on language learners' motivation, findings have been mixed, and most have not examined (or measured) motivation longitudinally, indicating a need for more research that tracks learners' motivation over time to better understand how it may change with chatbot use over time. This is particularly important given the potential of technology to enhance learner motivation (Reinders et al., 2022). Second, most research on chatbots focused on their potential as language practice tools and less attention has been paid to the potential of chatbots to facilitate SRL-related interactions. Finally, the common critique of traditional chatbots' limitations (Huang et al., 2022) suggests a need to empirically investigate whether the advanced capabilities of generative AI chatbots such as ChatGPT can address these limitations and provide a more effective and engaging learning experience (Kohnke et al., 2023).

3 The Present Study

The present study aims to address the gaps identified above by examining the effects of generative AI on students' language learning motivation over an extended period of time. We developed SRL Chatbots that focus on promoting SRL skills by asking learners to reflect on their learning. The study was conducted in two phases. In Phase 1, the chatbot was implemented using traditional, pre-scripted chat technology with fixed SRL-related questions; In Phase 2, a generative AI-powered chatbot was used (described below in more detail). By measuring motivation on a weekly basis over 15 weeks, we sought to capture the motivational dynamics more precisely and examine how it changed under these conditions. Our overarching research question was: what are the trajectories of students' learning motivation in the pre-scripted and generative AI phases?

4 Methodology

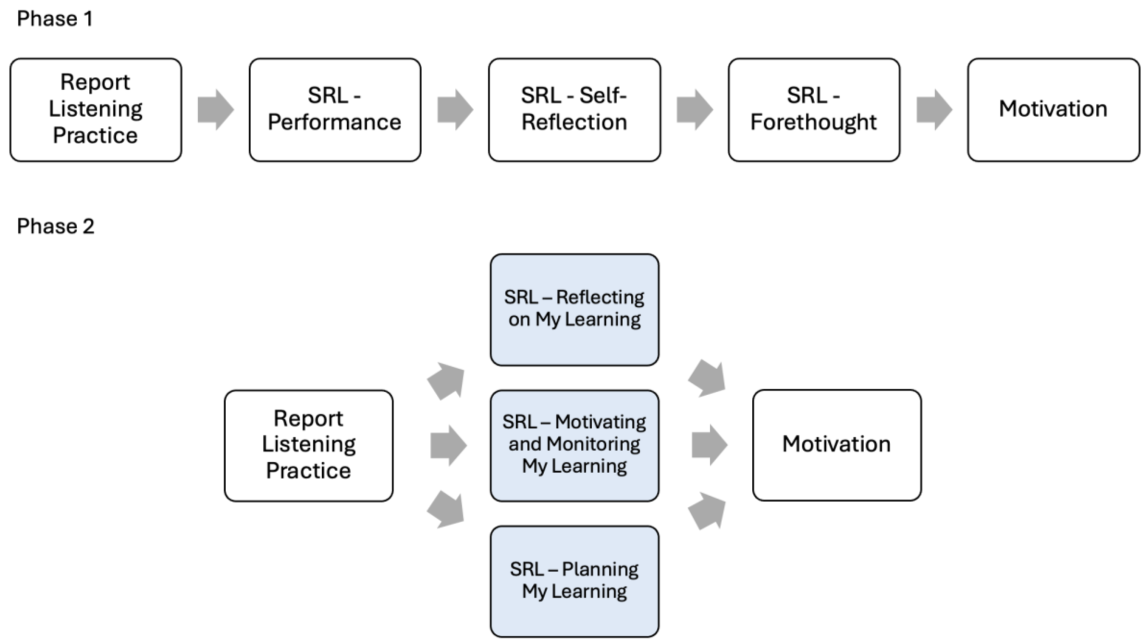
4.1 Participants and the Teaching Context

The participants in this study were 24 (n_{female} = 13) first-year English major students at a first-tier university in northern China (M_{age} = 18.3; SD = 0.49). All participants were Chinese and were enrolled in a compulsory English listening course designed for first-year English major students, with an average proficiency level of CEFR B2. The course was compulsory, with two credit hours of class time each week. It aimed to develop students' ability to understand a variety of listening materials, such as academic lectures, campus life conversations, and news broadcasts. The course placed a strong emphasis on fostering autonomy and encouraging students to take responsibility for their own learning. As part of the course requirements, students were expected to engage in independent listening practice outside of class, accounting for 20% of their final grade. The purpose was to provide students with a structure to explore and develop their listening skills on their own. Students were encouraged to engage with a variety of listening materials and accents based on their individual needs and interests, and to report their practice, including time spent, materials used, and accents encountered, in an interactive chatbot-powered learning journal. The students were informed at the beginning of the course that the assessment of this assignment would prioritize regular engagement and exposure to diverse listening materials.

4.2 Chatbot Design

The chatbot was designed using Voiceflow, an AI agent building platform. The platform has a visual interface that allows for complex conversation flows to be set up with little or no programming. Figure 1 is an overview of the conversation flows in the two phases of this study.

Figure 1 Overview of Conversation Flows



Note. Boxes shaded in blue represent where generative AI was integrated.

4.2.1 Phase 1 Pre-scripted Chatbot

In Phase 1, the chatbot was set to ask the participants a set of fixed questions (see Figure 2 for screenshots of a mock conversation). The participants were first asked to provide their student ID, report the total length of time they practiced listening, and the source and accent of the materials they listened to in the past week (i.e., the "reporting listening practice" in Phase 1, Figure 1; Panel 1-3, Figure 2). They were then asked three questions in sequence, each corresponding to a phase of Zimmerman (2000)'s SRL model – performance, self-reflection, and forethought. Table 1 shows the pre-scripted questions and their corresponding SRL categories as specified in Clear and Zimmerman (2004).

Table 1 Pre-scripted Questions and Corresponding SRL Categories

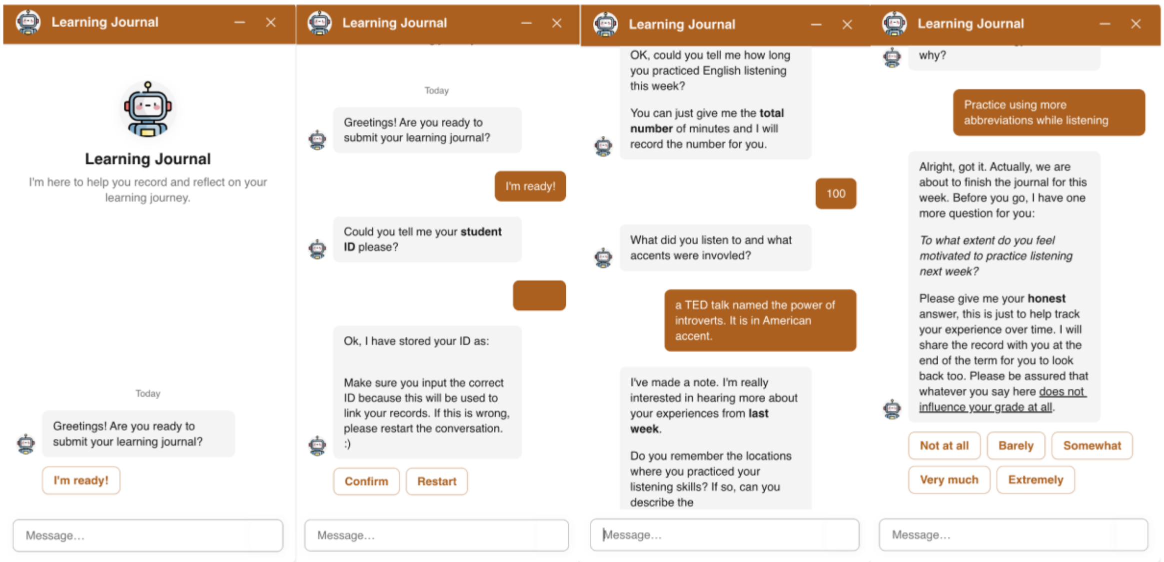
Categories in Cleary & Zimmerman (2004)		Pre-scripted Questions*
SRL Phases	Processes	

SRL – Performance	Attention Focusing	• I'm curious to learn about your experience of learning in the past week. Did you have to try to motivate yourself when practicing listening? If yes, what did you do? • I'm curious to know more about your experience last week. Were there any moments when you didn't feel up to doing the listening practice? If so, how did you handle those situations?
	Self-Recording	• I'm really interested in hearing more about your experiences from last week. Did you keep track of the locations where you practiced your listening skills? If so, can you describe the surroundings/environments of these places where you learned? • I'd love to hear more about last week's experience. Did you keep track of how long you studied English in total? If so, how long?
SRL – Self-Reflection	Self-Evaluation	• I wonder if you think you did well last week in terms of English listening? How did you determine whether you did well?
	Satisfaction	• I wonder how satisfied are you with your learning of listening last week? Why?
	Causal Attributions	• I wonder how well you think you did in last week's learning? What is the main reason for this?
	Adaptive Inferences	• I wonder how well you think you did in terms of English listening in the past week? What do you need to do to improve?
SRL – Forethought	Goal Setting	• Now, let's move on to look forward. Do you have a goal for your learning next week? Could you explain to me what you think? • Alright, let's move on to think about next week. Do you have a goal you want to achieve next week? Could you share your thoughts with me?
	Strategy Choice	• Let's move on to talk about the future. Do you have a learning strategy/method you would like to use next week? How do you decide what strategy to use and why?
	Self-Efficacy	• OK, let's move on to think about the future. How confident or certain are you that you can do well next week? • Got it, let's proceed to talk about next week. How confident are you that you can handle the learning

		problems/challenges in the coming week? I’m curious to hear your thoughts!
	Intrinsic Interest	• Noted. I wonder to what extent is practicing listening an interesting activity to you? Let me know how you feel.

*Each time, one question was randomly selected from the pool of questions within each of the three SRL phases (in the first column).

Figure 2 Screenshots of a Mock Conversation in Phase 1



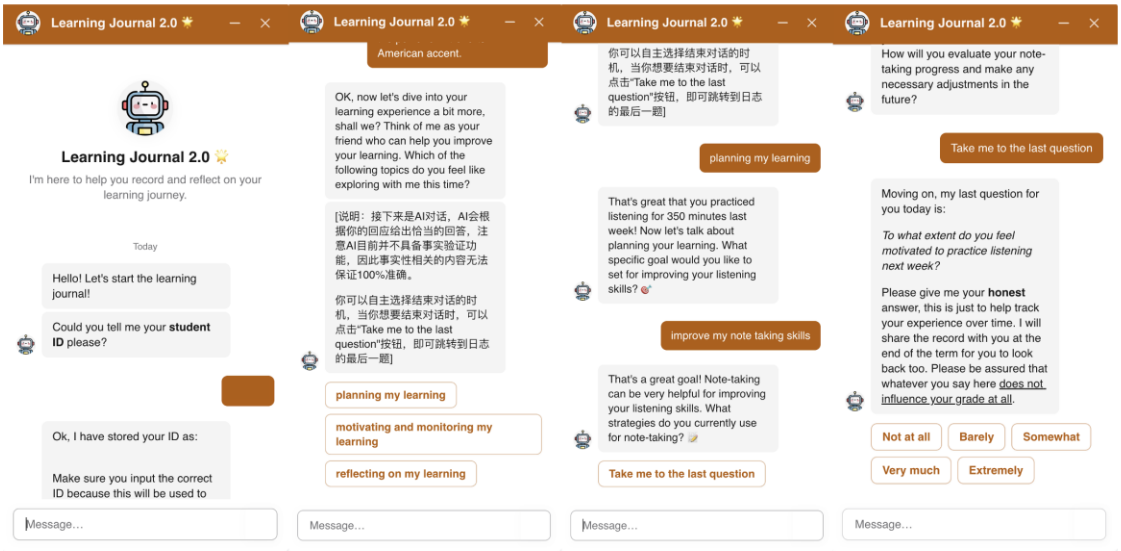
At the end of the conversation, there was a fixed question asking the participants to rate the extent to which they felt motivated to practice listening in the next week, with five options ranging from “Not at all” to “Extremely” (see Panel 4 in Figure 2).

4.2.2 Phase 2 Generative AI Chatbot

Phase 2 followed a similar structure to Phase 1. Students were first asked to report their practice, and at the end of the conversation, they were asked to rate their motivation (Panels 1 & 4, Figure 3). In the middle part of the conversation, students were given the option to engage in conversation with the AI regarding one of three topics – planning my learning, motivating and monitoring my learning, and reflecting on my learning. The specific behavior of the chatbot was engineered through customized prompts, which are texts given to large language models as input to guide the generation of responses. The chatbot was set with a system prompt to serve as a learning coach to help students with SRL. Each topic-specific bot also had their own prompt to initiate the AI conversation, to ensure the conversation was on-topic (or deviated as little as it can) throughout the conversation. The specific prompts used can be found in our online supplementary material (OSF link for blind review: https://osf.io/m8prf/?view_only=e9838488988a4a2a8b830b1a2b214c1b).

The three SRL topics were presented as parallel branches rather than sequential steps to ensure that the time on task could be more comparable to Phase 1, as a sequential design would have significantly increased the conversation length. A message in Chinese reminded the learner that this part would be powered by AI, together with a disclaimer that AI does not have fact-checking capabilities so the content may not be 100% accurate (Panel 3, Figure 3). Once the students clicked on a button, generative AI was activated. During the generative AI powered interaction, a button labeled “Take me to the last question” was always available to the student (Panel 3, Figure 3), allowing them to answer the final question (i.e., rating their motivation) and end the conversation at any time. Students were informed by the teacher that they needed to complete the last question to conclude the journal entry.

Figure 3 Screenshots of a Mock Conversation in Phase 2



Note. The Chinese in Panel 2 reads: "Note: The following conversation will be powered by AI, which will provide appropriate responses based on your input. Please be aware that the AI currently does not have the capability to verify facts, so the accuracy of factual information cannot be guaranteed 100%.

You may choose to end the conversation at any time. When you wish to conclude the dialogue, simply click on the 'Take me to the last question' button, and you will be directed to the final question of the journal."

4.3 Data Collection

Our study was comprised of two phases, with Phase 1 spanning from Week 1 – 8 and Phase 2 from Week 9 – 15. Participants were informed at the beginning of the semester of the aim of the learning journal (i.e., to reflect on their learning experience and report their listening practice to the teacher; *cf*. Section 4.1). In Week 9 they were informed that the chatbot would change to an AI-powered one.

The primary outcome measure for data analysis was participants' self-reported motivation, which was assessed at the end of each conversation using the question: "To what extent do you feel motivated to practice listening next week?" Responses were recorded on a five-point Likert scale (0 = "Not at all", 1 = "Barely", 2 = "Somewhat", 3 = "Very much", 4 = "Extremely").

4.4 Data Analysis

To examine the effect of generative AI on learners' motivational trajectories, piecewise growth curve models (also known as piecewise mixed effects models) were adopted. Specifically, this type of model allows for both random intercepts (i.e., initial values of motivation can vary) and random slopes (i.e., rate of motivational change can vary). Two splines (segments) with one knot fixed at Week 9 (when generative AI was introduced) were included in the models. A sequential model building process was adopted: a null model without any independent variables was fit first to decompose the variance in motivation into between-person and within-person variances, with intra-class correlation coefficient (ICC) values being calculated to indicate the degree of between-person variability present in the data. Then a random intercept model with fixed effects of time and a random slope model with random effects of time were fit. Likelihood ratio tests (with models refit with maximum likelihood) were performed to formally compare the model fit. Commonly used information criterion indices were calculated, namely Akaike information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978). Lower values for the AIC and BIC indicate better model fit. The model with the best fit was retained as the final model. This process was applied to the Content Chat group data and the SRL Chat group data separately.

All analyses were performed with R 4.2.1(R Core Team, 2022), using the piecewise growth curve models being fit with the *nlme* package. In support of methodological transparency, our R script and its corresponding output is shared on IRIS and OSF.

5 Findings

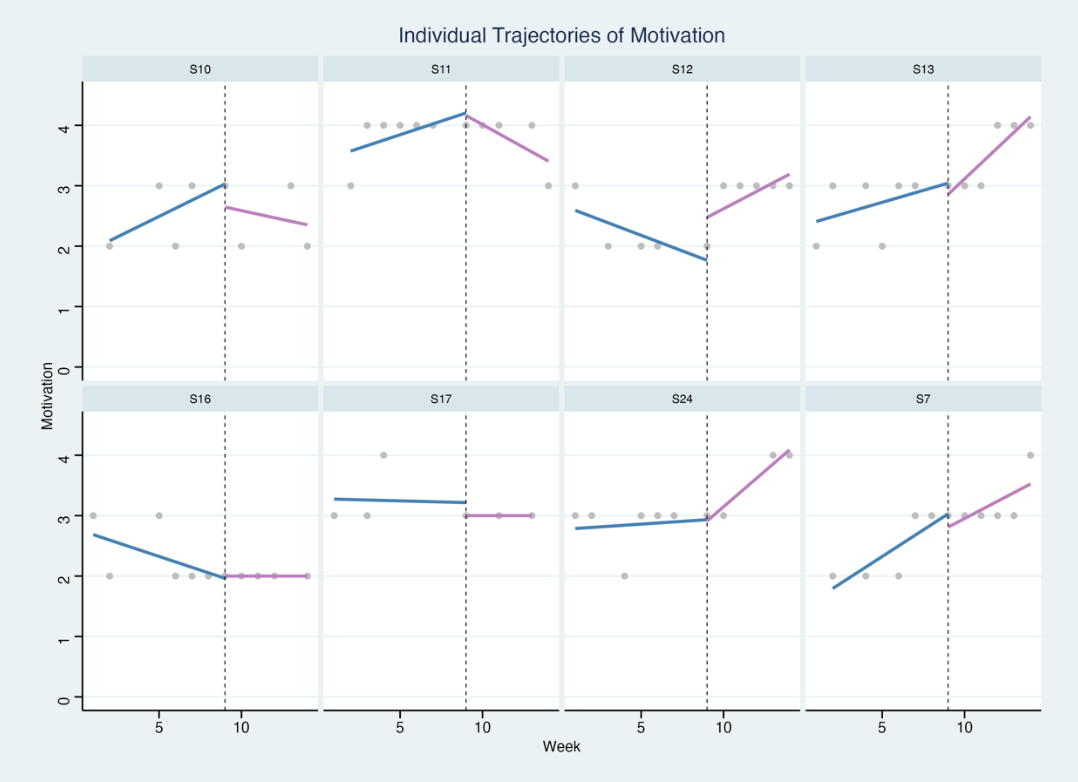
5.1 Descriptive Statistics

The descriptive statistics presented in Table 2 provide an overview of the motivation levels, response length (number of words) and number of turns (a pair of chatbot and student messages). It can be observed that motivation levels fluctuated over time, with the average in Phase 2 being slightly higher than in Phase 1. In terms of response length in words, it can be seen that the average in Phase 1 is slightly higher than in Phase 2, while the standard deviation in Phase 2 is greater. The number of turns in Phase 1 is very stable (participants either had 7 turns or 8 turns depending on whether they responded to the final standard message saying the journal has been completed). In contrast, there was much greater variability in Phase 2, with the average number of turns fluctuating over time and with greater variation between participants in each week. Figure 4 presents the motivational trajectories of a random sample of the data. Again, there is a noticeable degree of inter-individual differences and intra-individual variability.

Table 2 Descriptive Statistics (Week 1 – 15)

		Motivation	Response Length	Turns
Phase	Week	M (SD)	M (SD)	M (SD)
1	1	2.94 (0.54)	59.06 (28.91)	7.00 (0.00)
	2	2.61 (0.61)	58.33 (31.13)	7.17 (0.38)
	3	3.00 (0.76)	54.88 (29.23)	7.12 (0.35)
	4	3.00 (0.95)	56.50 (34.79)	7.08 (0.29)
	5	2.75 (0.91)	46.45 (27.27)	7.15 (0.37)
	6	2.72 (0.83)	49.06 (24.25)	7.22 (0.43)
	7	2.83 (0.79)	52.67 (28.32)	7.17 (0.38)
	8	3.11 (0.60)	51.78 (29.76)	7.22 (0.44)
	Average	2.83 (0.76)	53.36 (28.48)	7.14 (0.35)
2	9	2.62 (0.80)	46.86 (54.56)	7.71 (5.17)
	10	2.83 (0.79)	53.72 (67.84)	7.17 (4.93)
	11	3.00 (0.71)	37.76 (28.54)	6.12 (1.90)
	12	2.67 (0.89)	38.50 (27.12)	6.33 (3.11)
	13	3.06 (0.90)	52.88 (69.40)	7.29 (4.74)
	14	3.17 (0.92)	49.06 (53.98)	6.72 (3.21)
	15	3.50 (0.55)	65.17 (70.30)	9.00 (6.10)
	Average	2.93 (0.84)	47.96 (54.08)	7.06 (4.18)

Figure 4 Sampled Individual Trajectories of Motivation



5.2 Model Results

To assess the degree of between-person differences in motivational change, a null model (Model 1) was fit (Table 3). The intraclass correlation coefficient is .53, indicating that between-person variance explained 53% of the variance in motivation. Model comparison results can be found in Table 3, with Model 3 having the best fit and retained as the final model.

Table 3 Model Comparison Results

Model	*df*	AIC	BIC	LL	LRT test	χ^2	*p*
1	3	440.24	450.55	-217.12			
2	5	428.31	445.50	-209.16	1 vs. 2	15.93	.0003
3	10	410.06	444.44	-195.03	2 vs. 3	28.25	<.0001

Note. LL = Log Likelihood; LRT = Likelihood Ratio Test

Table 4 Final Model Results

Fixed Effects

	β	SE	95% CI	t	*p*
(Intercept)	2.83	0.11	[2.62, 3.05]	25.95	<.0001
Phase 1	-0.01	0.02	[-0.05, 0.02]	-0.66	.51
Phase 2	0.08	0.03	[0.02, 0.14]	2.66	.009
Random Effects					
			SD	Correlation (Intercept)	Phase 1
Participant (Intercept)			0.40		
Phase 1			0.06	.44	
Phase 2			0.10	-.75	.21
Model Fit					
R^2			Marginal	Conditional	
			0.03	0.67	

Note. Number of observations = 230; Number of groups = 24.

Figure 5 Model Implied Motivational Trajectories

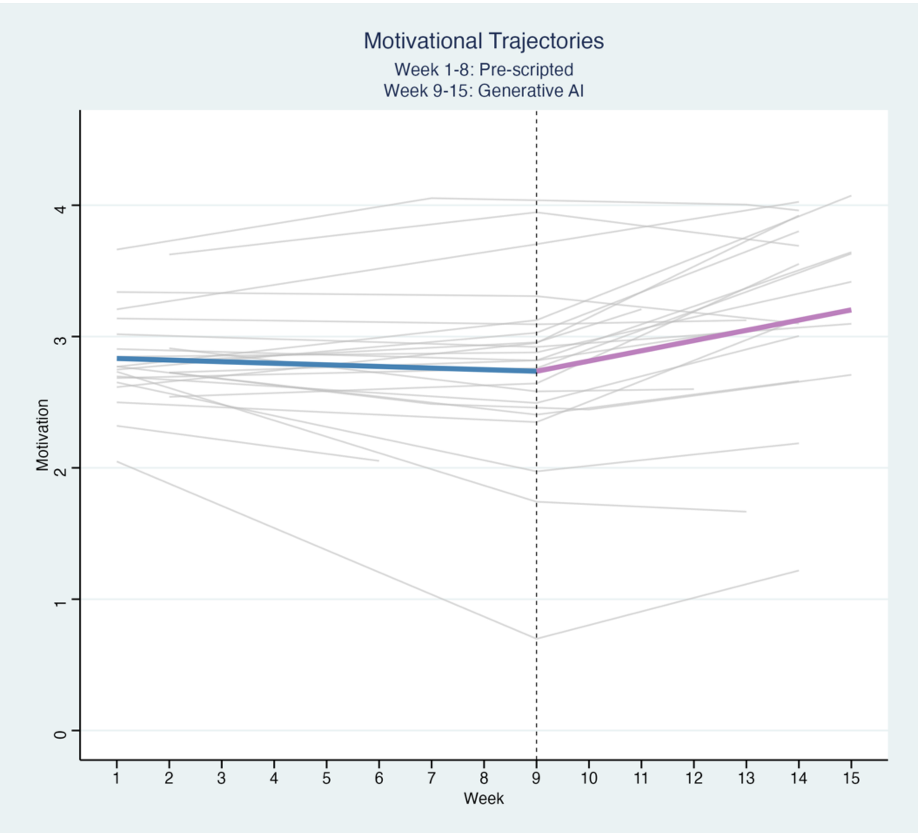


Table 4 shows the results of the final model. The fixed effects results showed that in Phase 1, there was a non-significant change in motivation (β=-0.01, 95% CI [-0.05, 0.02], p = .51), in the pre-scripted chat condition. In Phase 2, there was a significant positive change in motivation (β=0.08, 95% CI [0.02, 0.14], p = .009). This indicates that there is a slight but generally positive motivational growth during the second phase, i.e., in the generative AI chat condition (Figure 5).

The random effects results suggest there was variability in both the starting point and the rate of change among participants. The correlations among the random effects provide additional insights into the relationships between the intercepts and slopes at the participant level. The correlation between the participant intercept and the Phase 1 slope was .44, suggesting that participants with higher initial levels of motivation tended to have steeper positive slopes (or less negative slopes) during Phase 1. The correlation between the participant intercept and the Phase 2 slope was -.75, indicating that participants with higher initial motivation experienced a smaller improvement or even a decrease during the second phase. Notably, the correlation between the Phase 1 and Phase 2 slopes was .21, suggesting that there was weak association between the change patterns in the two phases at the participant level.

The conditional R^2 suggests the model explained a substantial amount of variation (67%) while fixed effects alone only accounted for 3% of the variance, which again, points to the substantial presence of individual differences in the data.

6 Discussion

Our study set out to investigate the impact of using SRL chatbots on learners' motivation in a two-phase study. Piecewise mixed effects model results show an upward trend in motivation after the introduction of generative AI. As one of the first studies that explored the use of state-of-the-art generative AI models in a long-term intervention, our findings supply some preliminary evidence on generative AI's superiority over traditional rule-based chatbot and its potential in enhancing students' learning motivation.

6.1 Pre-scripted vs. Generative AI Chatbot

The model results showed a stead trend of motivation in the pre-scripted phase. The absence of significant motivational increase can be attributed to the chatbot's limited capacity to provide support for learners' SRL needs, which aligns with previous research using traditional chatbots (Du et al., 2021; Hew et al., 2023). The abstract nature of SRL reflections, in conjunction with the absence of immediate, personalized feedback, might have diminished learners' interest in engaging with these cognitively demanding reflections over time (Hew et al., 2023).

According to SDT, autonomy refers to the sense of volition and self-endorsement of one's actions (Ryan & Deci, 2017; Vansteenkiste et al., 2020). When the learners in our study engaged with the pre-scripted chatbot, they may feel constrained by the predetermined conversation paths, which in turn would decrease their sense of volition and choice in the process. In other words, the setup of the pre-scripted chatbot may have thwarted the learners' need for autonomy due to its lack of flexibility and adaptability in responding to the learners.

In contrast, the upward trajectory of motivation in the generative AI phase suggests that generative AI may be particularly effective for supporting basic psychological needs (Ryan & Deci, 2017). As configured in the prompts (*cf.* Section 4.2.2), the generative AI model was instructed to converse with the learner in a friendly and constructive manner (including using emojis to make the conversation more engaging). An autonomy-supportive environment was thereby set up to allow for the learner to engage in more elaborated SRL-related thinking and reflection while maintaining a certain level of scaffolded guidance and structure, which has been found to be conducive to autonomous motivation in previous research (Oga-Baldwin & Nakata, 2015). This would account for why learners were more likely to experience a motivational growth in the generative AI phase. Alternatively, one might argue that a confounding factor may be the novelty effect of generative AI, akin to what was observed in previous research on the introduction of chatbots (Fryer et al. 2017, 2019). Notwithstanding its possibility, such effect would be less likely – given the duration of our study and the intensity of our measurement of motivation spanning over 7 weeks, before which the students were already exposed to chatbots for 8 weeks. Still, future research is needed to further extend the duration to truly disentangle short-term and long-term change patterns.

6.2 Interactivity as an AI affordance for Motivation

As an affordance of AI for motivation, interactivity (Zhang et al., 2023) provides another lens through which the contrasting results between the pre-scripted and generative AI chatbots can be explained. As a technological affordance, interactivity is characterized by reciprocity, responsiveness, nonverbal information, and speed of response (Johnson et al., 2006). Put simply, interactivity can be understood as the extent to which the interaction is perceived as "reciprocal, relevant, speedy, and characterized by the use of nonverbal information" (Johnson et al., 2006, p. 41). An information system with a high degree of interactivity has been found to have positive outcomes such as enhanced engagement, positive attitudes, and experience of flow (e.g., Guo et al. 2016). In this sense, the generative AI chatbot afforded a much higher level of interactivity compared to the pre-scripted chatbot – the immediate, responsive, and engaging responses with occasional emojis may have evoked learners' interest, positive emotions, and engagement.

6.3 AI and Heterogeneity in Intervention Effects

Our findings regarding the random effects in the model provide insights into the extent to which the intervention effects differed at the participant level. There was significant variability in participants' initial levels of motivation and their rates of change during each phase. One plausible explanation for the negative correlation between learners' initial motivation and Phase 2 change rate is the ceiling effect, which refers to the situation where individuals who started at a higher level enjoyed smaller gains due to limited room for improvement. Still, the weak correlation between the change patterns in the two phases suggests that generative AI had a transformative effect on learners' motivation (*cf.* the individual trajectories in Figure 5).

Our findings serve as an example of how individual differences in motivation can manifest in technology-enhanced learning environments. The heterogenous trajectories observed in our study aligns with Jeon (2024), which identified a range of interacting influences on students' motivation, ranging from learners' perceived competence to the chatbots' pedagogical value. The concept of motivational profiles or learner profiles (e.g., Csizér & Dörnyei, 2005; Liu & Oga-Baldwin, 2022; Oga-Baldwin & Fryer, 2016) is particularly relevant in this context. Identifying and monitoring learners' profiles would be crucial for customizing AI chatbots in a way that best caters to learners' needs.

AI is likely to further bring individual differences to the fore, due to its capabilities for personalization and adaptivity. As AI advances, it may not only accommodate but also amplify these differences, leading to a broader spectrum of learning experiences and outcomes, which our findings seem to corroborate. As indicated by the substantial variance of change patterns in the two phases of our study, generative AI may interact with individual differences in convoluted ways and potentially contribute to more divergent learning outcomes. This indicates new challenges that come with the opportunities brought by generative AI, which researchers and practitioners may need to grapple with in the years to come.

6.4 Limitations and future directions

While our findings supply some initial evidence of the effects of generative AI on motivation in chatbot-assisted language learning, it is essential to acknowledge some limitations. First, the small sample size from a single educational setting over one semester limits the generalizability of the findings. Future research should replicate and extend this study with larger and more diverse samples, longer study durations, and more measurement points to allow for formal testing of differences between phases, for example. Second, the study did not investigate the specific causes of heterogeneity in intervention effects. Given the high degree of between-person variability, future research could address this by measuring various individual difference factors to better understand the sources of variability in learners' responses and complement this with qualitative data, for example, of the unique ways learners interact with the chatbot. Additionally, this study only focuses on AI chatbot's impact on motivational outcomes, which captures a single (though crucial) facet of a complex variety of learning outcomes. Future studies could explore the effects of generative AI powered chatbots on other aspects of language learning, such as the effects on different language skills as well as specific SRL skills.

Conclusion

In our two-phase study, we examined the impact of chatbots on learners' motivation. Results showed a steady motivation trajectory with the pre-scripted chatbot, but an upward trajectory with the generative AI chatbot. Generative AI's advanced language capabilities and interactivity likely better supported learners' autonomy and competence needs. However, significant variability in individual intervention effects suggests AI may also amplify the differences in learning experiences and outcomes, which calls for more empirical research to disentangle the complex interactions between individual differences and AI-assisted language learning.

References

Al-Hoorie, A. H., Oga-Baldwin, W. L. Q., Hiver, P., & Vitta, J. P. (2022). Self-determination mini-theories in second language learning: A systematic review of three decades of research. *Language Teaching Research*, 36.

Ardasheva, Y., Wang, Z., Adesope, O. O., & Valentine, J. C. (2017). Exploring effectiveness and moderators of language learning strategy instruction on second language and self-regulated learning outcomes. *Review of Educational Research*, *87*(3), 544–582. https://doi.org/10.3102/0034654316689135

Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, *29*, 29–57.

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, *32*(8), 827–877.

Boekaerts, M., Zeidner, M., & Pintrich, P. R. (1999). *Handbook of self-regulation*. Elsevier.

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, *41*(5), 537–550. https://doi.org/10.1002/pits.10177

Csizér, K., & Dörnyei, Z. (2005). Language learners' motivational profiles and their motivated learning behavior. *Language Learning*, *55*(4), 613–659. https://doi.org/10.1111/j.0023-8333.2005.00319.x

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Taylor & Francis Group.

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.

Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.). (2015). *Motivational dynamics in language learning*. Multilingual Matters.

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation* (Third edition). Routledge.

Du, J., Huang, W., & Hew, K. F. (2021). Supporting students goal setting process using chatbot: Implementation in a fully online course. *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, 35–41. https://doi.org/10.1109/TALE52509.2021.9678564

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Gallacher, A., Thompson, A., Howarth, M., Taalas, P., Jalkanen, J., Bradley, L., & Thouësny, S. (2018). "My robot is an idiot!"–Students' perceptions of AI in the L2 classroom. *Future-Proof CALL: Language Learning as Exploration and Encounters–Short Papers from EUROCALL*, 70–76.

Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *The Journal of Information and Systems in Education*, *13*(1), 1–7.

Han, Z. (2024). Chatgpt in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition*, 1–6. https://doi.org/10.1017/S0272263124000111

Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education*, *35*(1), 40–68. https://doi.org/10.1007/s12528-022-09338-x

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot- supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1–2), 1–26.

Joe, H.-K., Hiver, P., & Al-Hoorie, A. H. (2017). Classroom social climate, self-determined motivation, willingness to communicate, and achievement: A study of structural relationships in instructed second language settings. *Learning and Individual Differences*, *53*, 133–144. https://doi.org/10.1016/j.lindif.2016.11.005

Johnson, G. J., Bruner Ii, G. C., & Kumar, A. (2006). Interactivity and its facets revisited: Theory and empirical test. *Journal of Advertising*, *35*(4), 35–52. https://doi.org/10.2753/JOA0091-3367350403

Kato, S., & Mynard, J. (2015). *Reflective dialogue: Advising in language learning*. Routledge.

Kim, N.-Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning*, *19*(4).

Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, *22*(3).

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of Robot-Assisted Language Learning. *ReCALL*, *23*(1), 25–58. https://doi.org/10.1017/S0958344010000273

Liu, M. (2024). Mapping the landscape of research on the L2 motivational self system theory (2005–2021): A bibliometric and text network analysis. *System*, *120*, 103180. https://doi.org/10.1016/j.system.2023.103180

Liu, M., & Oga-Baldwin, W. L. Q. (2022). Motivational profiles of learners of multiple foreign languages: A self-determination theory perspective. *System*, *106*, 102762. https://doi.org/10.1016/j.system.2022.102762

Oga-Baldwin, W. L. Q., & Fryer, L. K. (2016). *Exploring motivational profiles in public elementary school English classes*. 7.

Oga- Baldwin, W. L. Q., & Nakata, Y. (2015). Structure also supports autonomy: Measuring and defining autonomy- supportive teaching in Japanese elementary foreign language classes. *Japanese Psychological Research*, *57*(3), 167–179. https://doi.org/10.1111/jpr.12077

Pintrich, P. R. (1995). Understanding self- regulated learning. *New Directions for Teaching and Learning*, *1995*(63), 3–12.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451–502). Elsevier.

R Core Team. (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna*. https://www.R-project.org/.

Reinders, H., Dudeney, G., & Lamb, M. (2022). *Using technology to motivate learners*. Oxford University Press.

Reinders, H., & Hubbard, P. (2013). CALL and learner autonomy: Affordances and constraints. *Contemporary Computer Assisted Language Learning*, 359–375.

Rose, H., Briggs, J. G., Boggs, J. A., Sergio, L., & Ivanova-Slavianskaia, N. (2018). A systematic review of language learner strategy research in the face of self-regulation. *System*, *72*, 151–163.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

Vansteenkiste, M., Ryan, R. M., & Soenens, B. (2020). Basic psychological need theory: Advancements, critical themes, and future directions. *Motivation and Emotion*, *44*(1), 1–31. https://doi.org/10.1007/s11031-019-09818-1

Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE—Virtual Immersive Language Learning and Gaming Environment: Immersion and presence. *British Journal of Educational Technology*, *48*(2), 431–450.

Xu, Y., & Warschauer, M. (2020). *Exploring young children's engagement in joint reading with a conversational agent*. 216–228.

Zhang, J., Liu, Z., Lv, H., & Jiang, M. (2023). Ai in e-learning: The affordance perspective. *Behaviour & Information Technology*, 1–30. https://doi.org/10.1080/0144929X.2023.2287660

Zimmerman, B. J. (2000). Attaining Self-Regulation. In *Handbook of Self-Regulation* (pp. 13–39). Elsevier. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American*

Educational Research Journal, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, *48*(3), 135–147.

Zou, B., Reinders, H., Thomas, M., & Barr, D. (2023). Editorial: Using artificial intelligence technology for language learning. *Frontiers in Psychology*, *14*, 1287667. https://doi.org/10.3389/fpsyg.2023.1287667