HEC MONTRÉAL

Impact of Feedback Timing on Metacognition

par

Asikaer Nadila

Sylvain Sénécal

HEC Montréal

Codirecteur de recherche

Constantinos K. Coursaris

HEC Montréal

Codirecteur de recherche

Sciences de la gestion

(User Experience)

Mémoire présenté en vue de l'obtention

du grade de maîtrise ès sciences en gestion

(M. Sc.)

Dec 2024

HEC MONTRĒAL

Comité d'éthique de la recherche

ATTESTATION D'APPROBATION ÉTHIQUE COMPLÉTÉE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet des approbations en matière d'éthique de la recherche avec des êtres humains nécessaires selon les exigences de HEC Montréal.

La période de validité du certificat d'approbation éthique émis pour ce projet est maintenant terminée. Si vous devez reprendre contact avec les participants ou reprendre une collecte de données pour ce projet, la certification éthique doit être réactivée préalablement. Vous devez alors prendre contact avec le secrétariat du CER de HEC Montréal.

Nom de l'étudiant(e) : Asikaer Nadila

Titre du projet supervisé/mémoire/thèse : AI-Driven Feedback Timing and Its Impact on Learners' Metacognitive Judgments

Titre du projet sur le certificat : Etude de l'impact de la rétrocation d''un agent conversationnel à base vocale (VCA) sur les expériences d'apprentissage des langues et sur leurs résultats

Projet # : 2024-5921

Chercheur principal / directeur de recherche : Sylvain Sénécal

Cochercheurs : Pierre-Majorique Léger; Constantinos K. Coursaris; Marc Fredette; Jared Boasen; Alexander John Karran; Frédérique Bouvier; David Brieugne; Luis Carlos Castiblanco; Juan Fernandez Shaw; Salima Tazi; Xavier Côté; Shang Lin Chen; Élise Imbeault; Elham Rashidi Ranjbar; Asikaer Nadila

Date d'approbation initiale du projet : 31 mai 2024

Date de fermeture de l'approbation éthique pour l'étudiant(e) : 25 octobre 2024

Maurice Lemelin
Président
CER de HEC Montréal

Signé le 2024-10-25 à 11:04

Résumé

Cette recherche examine l'impact du moment de la rétroaction sur la métacognition dans le contexte de l'apprentissage des langues facilité par l'IA conversationnelle. L'étude comble une lacune dans la littérature, étant donné que la plupart des travaux antérieurs sur la rétroaction portaient principalement sur l'interaction humain-humain, négligeant lles contextes d'interaction humain-AI. L'objectif de la recherche est d'explorer comment le moment de la rétraction (feedback timing) affecte les états émotionnels et cognitifs (valence, arousal, charge cognitive) et comment ces états médient la relation entre le le moment de rétroaction et le jugement métacognitif. L'étude porte sur 30 adultes anglophones ayant un niveau de compétence linguistique A2 en français, assignés de manière aléatoire à des conditions de rétroaction immédiat ou différé. La recherche utilise un modèle inter-sujets et les mesures sont obtenues par des échelles d'auto-évaluation et des mesures physiologiques. Les résultats indiquent que le moment de rétroaction n'a pas d'impact sur la valence, l'excitation ou la charge cognitive et que les deux conditions ne diffèrent pas statistiquement en ce qui concerne la précision du jugement métacognitif. Cependant, une valence perçue plus élevée était associée à une plus grande précision métacognitive en matière d'intelligibilité, tandis qu'une charge cognitive accrue était associée à une meilleure précision dans les jugements métacognitifs relatifs à l'accentuation. Cela signifie que, bien que le moment du feedback n'ait pas d'effet direct sur les résultats métacognitifs, la valence émotionnelle ainsi que la charge cognitive ont des effets significatifssur l'amélioration de la précision métacognitive.

Mots clés: Métacognition, Rétroaction, Valence, Arousal, Charge cognitive, IA générative

Méthodes de recherche: Conception expérimentale entre sujets pour évaluer l'impact de la synchronisation du feedback sur la métacognition. L'état émotionnel et l'état cognitif, y compris la valence, l'éveil et la charge cognitive ont été mesurés à l'aide d'évaluations auto-rapportées et psychophysiologiques.

Abstract

The research examines the impact of feedback timing on metacognition in the context of AI-mediated language learning. The study addresses a gap in the literature, since most previous works were mostly centered on human-to-human interactions, overlooking the influence of AI-mediated feedback on metacognition. The objective of this research is to explore how feedback timing affects emotional and cognitive states (valence, arousal, cognitive load) and how these states mediate the relationship between feedback timing and metacognitive judgment. The study involves 30 adult English speakers at the A2 French language proficiency level. Participants are randomly assigned to either immediate or delayed feedback conditions during reading aloud task in French, where AI tutor provide pronunciation feedback. The research utilizes a between-subjects design, gathering data through self-report scales and physiological measures. The results indicate that feedback timing does not have an impact on valence, arousal or cognitive load. And the two conditions were not found to statistically differ in metacognitive judgment accuracy. However, more positively perceived valence was associated with higher metacognitive accuracy in comprehensibility, while increased cognitive load was associated with improved accuracy in metacognitive judgments for accentedness. This implies that, although feedback timing may not have a direct effect on metacognitive outcomes, emotional valence and cognitive load have a significant effect on improving metacognitive accuracy.

Keywords: Metacognition, Feedback, Valence, Arousal, Cognitive load, Generative AI

Research methods: Between subject experimental design to assess the impact of feedback timing on metacognition. Emotional state and cognitive state, including valence, arousal, and cognitive load were measured using self-reported and psychophysiological assessments.

Table of contents

List of tables and figures

List of tables

Chapter 1

Chapter 2

List of figures

Chapter 2

List of abbreviations and acronyms

AI: Artificial Intelligence

LLM: Large Language Model

Gen AI: Generative Artificial Intelligence

Preface

An authorization to write the following dissertation has been granted by the administrative direction of the Master of Science program specializing in user experience. This dissertation is written in the form of an article. The agreement of all co-authors for this article has been obtained, as shown in the appendix.

In November 2024, the HEC Montréal Research Ethics Board (CER) approved the research project (Certificate #2024-5921), as shown in page 2.

The article examines the impact of AI-mediated feedback timing on metacognition.

Acknowledgements

Firstly, I would like to thank my co-supervisors, Sylvain Sénécal and Constantinos K. Coursaris, for their support throughout my thesis project. I am grateful for the opportunity they gave me to work on a topic that interests me. I have learned so much from their guidance and expertise.

I am also incredibly thankful to all the Tech3Lab members and my project partner, Elham, who provided immense support during the data collection phase. I could not have come this far without their technical and emotional support. A special thanks goes to Shang Lin Chen—without him, I might still be lost in the maze of data analysis.

Lastly, I want to express my gratitude to my wonderful parents, Aziguli and Asikaer, who, despite knowing little about my thesis topic, always cheered me on during challenging moments and supported me financially.

Finally, I would like to thank myself. Thank you for not giving up, even when you thought about it 173,848,473,9 times. If you can finish writing a thesis, you can find a job too. The end of your academic journey does not mean the beginning of unemployment—it is just the start of a new chapter.

Chapter 1

Introduction

Context and Background

As AI drastically changes industries and daily life, its global user base reached over 250 million in 2023, more than double the figure from 2020, and is expected to exceed 700 million by 2030 (Statista, 2023). In the U.S., the Generative AI market is expected to grow from 16 billion dollars in 2023 to over 60 billion by 2030. This rapid growth is driven by tools such as ChatGPT and Midjourney that have captured public attention.

AI-driven voice technologies have emerged as powerful tools in educational contexts, offering new opportunities that are challenging traditional teaching methods. Unlike static e-learning technology, Gen AI provides real-time, adaptable, and personalized feedback. This is particularly beneficial to those who requires precise, responsive guidance to master a new language. By providing low-cost and adaptable solution, Gen AI can make advance tutoring available to a larger audience.

Yet, there are critical questions regarding their potential impact on the learning process.

Research Gap

Although there is research support that corrective feedback enhances verbal learning (Metcalfe & Kornell, 2007), behaviorists claim that to effectively learn, one must receive feedback directly after issuing a response because even slight delays significantly hinder acquisition (Saltzman, 1951). Studies have supported the idea that immediate feedback improves learning and retention (Kulik & Kulik, 1988), yet other research suggests delayed feedback may also enhance learning (Butler & Roediger, 2008). However, most of this research has focused on traditional educational contexts (Fu & Li, 2022; Rassaei, 2023; Corral et al., 2021), leaving a gap in understanding how feedback timing functions in AI-mediated environments. Feedback not only informs learners about their performance but also shapes how they evaluate their own learning (Luo & Liu, 2023; Haddara & Rahnev, 2022; Butler, Karpicke, & Roediger, 2008) - a process known as metacognitive judgment. This study examines how the timing of feedback (immediate or delayed) affects metacognitive judgments in AI-mediated language learning environments.

Purpose and Objectives

The primary objective of this research is to investigate the impact of feedback timing (immediate versus delayed) on metacognitive judgments within an AI-mediated language learning context. Thus, the main research question of this paper is:

RQ1: To what extent does the timing of feedback impact metacognitive judgment?

A secondary objective is to explore the relationship between emotional and cognitive states and metacognitive judgment accuracy during AI-mediated language learning sessions. While research has examined emotional and cognitive factors in feedback (Aghaei Pour et al., 2010;

Kluger, Lewinsohn, & Aiello, 1994; Moreno, 2004), few studies have done so within AI-based environments with a focus on the timing of feedback. This takes us to our secondary research question:

> RQ2: To what extent do learners' emotions and cognition play a role in the relationship between feedback timing and metacognitive judgment?

Significance of the Study

This study contributes theoretically by extending the understanding of how conversational AI affects metacognitive processes like self-monitoring and self-evaluation, filling a gap in the literature. Methodologically, the use of GPT-4 as a conversational AI tutor shows a new approach to delivering customized feedback on pronunciation, offering a replicable framework for future research. Practically, AI tutors can be used as a tool for learning second language, helping speech therapy patients, and young students struggling with reading or pronunciation. Additionally, addressing the research question will help educators and AI developers understand how feedback timing influences self-regulated learning, leading to more effective AI tutoring systems.

Theoretical Framework

This study builds on Nelson and Narens' Metacognitive Model (1990), which distinguishes between two levels of cognitive processing: the object-level, where tasks like learning occur, and the meta-level, where these tasks are evaluated. The model focuses on bidirectional information flow—monitoring (object to meta) and control (meta to object)—to regulate learning. The study examines how feedback timing (immediate versus delayed) influences monitoring processes. As well as exploring how AI-mediated feedback shapes metacognitive judgment, contributing to the understanding of self-regulated learning in technology-based contexts.

Method

This study used a between-subjects experimental design to examine the effects of feedback timing (immediate vs. delayed) on metacognitive judgments in an AI-mediated language learning context. Participants were assigned to either an immediate or delayed feedback condition. They completed reading-aloud tasks in French while interacting with an AI tutor providing pronunciation feedback. Throughout the sessions, emotional states (valence and arousal), cognitive load, and metacognitive judgments (comprehensibility and accentedness) were assessed. Statistical analyses were conducted to compare how feedback timing affected emotional and cognitive states, as well as the accuracy of metacognitive judgments, across the two groups

Scope

This study focuses on how the timing of AI-mediated feedback (immediate versus delayed) affects beginner-level French learners' metacognitive judgments during reading-aloud tasks. Two main limitations are recognized. First, the relatively short duration of the trials, with a limited number of sentences, may have restricted the observable differences between feedback condition. In future studies, extending the trial duration could better capture the difference between two groups. Secondly, the small sample size due to logistical reasons may have decreased the statistical power of the findings. Increasing the number of participants in the subsequent studies would enhance the validity and generalizability of the outcomes.

Thesis Structure

The thesis is structured into four main chapters, followed by a bibliography and appendix. Chapter 1 introduces the research context, objectives, and significance, focusing on how AI-mediated feedback timing influences metacognitive processes in language learning. Chapter 2 presents Article 1, providing a detailed examination of the study's methodology, data analysis, and findings related to feedback timing, metacognitive accuracy, emotional states, and cognitive load. Chapter 3 targets managers as an audience and deals with the methods of the AI tutor and their practical use. Chapter 4 synthesizes major findings, showing how the research adds to existing knowledge, specifying its limitations, and giving suggestions for further research. The bibliography consists of all references, and the appendix includes supplementary materials relevant to the study.

Personal Contribution

Since this thesis was conducted within the Tech3Lab, which includes multiple collaborators contributing at different stages and levels, Table 1 outlines my individual intellectual contributions across each aspect of the thesis. In accordance with lab standards, a minimum overall contribution of 50% is expected from the student. Any dimensions where my contributions exceed 50% indicate leadership and ownership of those respective phases.

Table 1: Contribution to the responsibilities of the research project phases

Research Activity	**Contribution**
Research Questions	Formulating appropriate research questions based on the research partner organization's expectations and needs – 60% *Support from the directors and supervisor was provided to determine the research partner's expectations and needs. *Support from the directors and supervisor was provided to formulate appropriate research questions.
Experimental Design	Conceiving and formalizing the experimental protocol – 50% *Members of the Tech3lab conceived the experimental protocol.
Stimuli	Creating the stimuli – 50% *The stimuli were co-created with a fellow student
Questionnaires	Creating the questionnaire on Qualtrics – 50% *Support from the Tech3lab members was provided
Ethics	Requesting ethical approval from CER - 50%
Pretests	Pre-test in the lab – 40% * Research assistants from Tech3lab were partially responsible for this portion.
Recruitment	Recruiting participant – 40%
Data Collection	Data collection in the lab – 50% *Research assistants from Tech3lab were partially responsible for this portion.
Analyzing	Analyzing data – 60% * Support from the lab's statistician was of great help in the analysis process.
Writing	Writing introduction, scientific and managerial articles – 90% *Support from the directors and supervisor was provided to guide and revise the articles.

References

Statista. (2023). *Generative Artificial Intelligence (AI)*. Retrieved October 3, 2024, from https://www.statista.com/study/133112/generative-artificial-intelligence-ai/

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In Psychology of learning and motivation (Vol. 26, pp. 125-173). Academic Press.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. Psychonomic Bulletin & Review, 14, 225-229.

Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. Studies in Second Language Acquisition, 44(1), 2-34.

Rassaei, E. (2023). The interplay between corrective feedback timing and foreign language anxiety in L2 development. Language Teaching Research, 13621688231195141.

Corral, D., Carpenter, S. K., & Clingan-Siverly, S. (2021). The effects of immediate versus delayed feedback on complex concept learning. Quarterly Journal of Experimental Psychology, 74(4), 786-799.

Saltzman, I. J. (1951). Delay of reward and human verbal learning. Journal of experimental psychology, 41(6), 437.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. Review of educational research, 58(1), 79-97.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Memory & cognition, 36(3), 604-616.

Luo, T., & Liu, C. (2023). The impact of feedback on metacognition: Enhancing in easy tasks, impeding in difficult ones. Consciousness and Cognition, 116, 103601.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 918.

Aghaei Pour, P., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. In V. Aleven, J. Kay, & J. Mostow (Eds.), Intelligent tutoring systems (Lecture Notes in Computer Science, Vol. 6094). Springer. https://doi.org/10.1007/978-3-642-13388-6_31

Kluger, A. N., Lewinsohn, S., & Aiello, J. R. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. Organizational Behavior and Human Decision Processes, 60(2), 276-299.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. Instructional science, 32(1), 99-113.

Chapter 2

The Impact of AI-Driven Feedback on Learners' Metacognitive Judgments

Highlights (3-5 bullet points)

- The timing of feedback (immediate or delayed) did not affect how accurately participants assessed their own learning.
- There is no difference between immediate and delayed feedback condition on participants' valence, arousal, and cognitive load.
- Higher emotional valence was associated with better self-assessment on how understandable their pronunciation was.

Abstract (250 words)

The research examines the impact of feedback timing on metacognition in the context of AI-mediated language learning. The study addresses a gap in the literature, since most previous works were mostly centered on human-to-human interactions, overlooking the influence of AI-mediated feedback on metacognition. The objective of this research is to explore how feedback timing affects emotional and cognitive states (valence, arousal, cognitive load) and how these states mediate the relationship between feedback timing and metacognitive judgment. The study involves 30 adult English speakers at the A2 French language proficiency level. Participants are randomly assigned to either immediate or delayed feedback conditions during reading aloud task in French, where AI tutor provide pronunciation feedback. The research utilizes a between-subjects design, gathering data through self-report scales and physiological measures. The results indicate that feedback timing does not have an impact on valence, arousal or cognitive load. And the two conditions were not found to statistically differ in metacognitive judgment accuracy. However, more positively perceived valence was associated with higher metacognitive accuracy in comprehensibility, while increased cognitive load was associated with improved accuracy in metacognitive judgments for accentedness. This implies that, although feedback timing may not have a direct effect on metacognitive outcomes, emotional valence and cognitive load have a significant effect on improving metacognitive accuracy.

Keywords: Metacognition, Feedback, Valence, Arousal, Cognitive load, Conservational AI

1. Introduction

As AI drastically changes industries and daily life, its global user base reached over 250 million in 2023, more than double the figure from 2020, and will go beyond 700 million by 2030 (Statista, 2023). In the U.S., the Generative AI market is expected to grow from 16 billion dollars in 2023 to over 60 billion by 2030. This rapid growth is driven by tools such as ChatGPT and Midjourney that have captured public attention. The advancement of technology, particularly the rise of voice-based conversational AI, has created new opportunities for learning experiences.

This brings us to perhaps a more critical question in a world where AI is revolutionizing learning: How might AI most effectively support not only what we learn but also how we think about our learning? The current challenge remains in making AI-driven learning platforms as effective as possible, particularly in enhancing metacognitive skill, which is learners' ability to reflect on, monitor, and regulate their own learning through feedback. Generative AI technologies, including large language models, create possibilities for interactive, real-time support of language learning. Currently, there are no solutions using LLMs, such as ChatGPT-

Voice, that act as tutors to provide timely and effective feedback while learners practice reading aloud. This gap significantly limits the potential of AI in language education.

Using LLMs for feedback in language learning is crucial because it provides an accessible, and cost-effective alternatives to traditional human tutors. Unlike other conventional e-learning methods that are less dynamic, AI-driven platforms provide flexible, immersive experiences with great potential to significantly enhance the level of learners' engagement and progress. (Taj et al., 2017). These systems can also provide immediate, adaptive feedback, bridging the gap between human tutoring and technology-based instruction. They encourage learner autonomy and possibly even a revolution in the process of language acquisition. (Chen et al., 2021).

Furthermore, one of the important aspects of learning is metacognition—the process of 'thinking about thinking'— which allows learners to self-regulate by planning, monitoring, and then evaluating their understanding and performance (Brown, 1987). In an era shaped by AI and digital learning, developing metacognitive skills has become more important than ever. For educational tools to be truly effective, not only does it need to communicate information but also foster the development of metacognitive skills (Azevedo & Aleven, 2013). Thus, studying how AI-mediated feedback influences metacognition can provide valuable insights into enhancing the effectiveness of educational technologies.

Most research on feedback in learning has focused on traditional classroom settings and human tutors, looking into its impacts on things like retention, comprehension, and performance (Hattie & Timperley, 2007; Smith & Kimball, 2010; Webb et al., 1997). These studies have shown how learners can effectively process new information and have provided a foundation for understanding the role of feedback in learning. Now, with the rise of AI in education, we're seeing voiced-based AI that makes learning more personalized, interactive, and engaging. These advancements open new ways to think about how feedback plays a role in learning.

There are two significant gaps in the current research. The first is understanding the role of AI-mediated feedback compares to human feedback. The second is exploring how feedback timing impacts metacognitive processes like self-monitoring, self-regulation, and self-evaluation. Studies have shown that immediate feedback improves learning and retention (Kulik & Kulik, 1988), while other research suggests delayed feedback may also enhance learning (Butler & Roediger, 2008). However, most of this research has focused on traditional educational contexts (Fu & Li, 2022; Rassaei, 2023; Corral et al., 2021), providing little insight as to how feedback timing works in AI-mediated environments. Feedback not only provides learners with insights about their performance but also shapes how they evaluate their own learning (Luo & Liu, 2023; Haddara & Rahnev, 2022; Butler, Karpicke, & Roediger, 2008). The difference between feedback provided by AI tutors versus human instructors, especially in terms of its effect on metacognitive judgment accuracy remains underexplored. Addressing these gaps is essential for creating AI-driven educational tools that enhance both learning and metacognitive skills.

This study examined how the timing of feedback (immediate versus delayed) affects metacognitive judgments in the context of AI-mediated language learning. Specifically, we focused on understanding how feedback timing influences learners' emotional states (valence and arousal), cognitive load, and the accuracy of their metacognitive judgments. The accuracy of the metacognitive judgment is defined as the extent to which an individual's self-assessment of their performance aligns with their actual performance. Accuracy in metacognition in our research has been implemented within the context of learners' French pronunciation, focusing on comprehensibility (ease of understanding), and accentedness (linguistic nativelikeness). Previous research has focused on the metacognitive process in human-to-human feedback, but the role of AI-mediated feedback has been largely overlooked. By integrating conversational AI, this research provides a new perspective on how technology can support metacognitive development and self-regulated learning.

This study investigated the effects of feedback timing on metacognitive judgments in a controlled experiment wherein participants completed reading tasks within an AI-mediated learning environment. Immediate and delayed feedback was provided to learners while measuring their emotional state (valence and arousal) and cognitive load. The main findings showed that the timing of feedback (immediate versus delayed) did not significantly impact learners' emotional states (valence and arousal), cognitive load, or the accuracy of their metacognitive judgments in comprehensibility and accentedness. However, the study found that higher perceived emotional valence was associated with higher metacognitive accuracy for comprehensibility and potentially for accentedness. Additionally, cognitive load was related to metacognitive judgment accuracy in accentedness but not in comprehensibility. These results suggest that while feedback timing may not directly impact metacognitive outcomes, emotional states and cognitive load play a significant role in improving metacognitive accuracy in AI-mediated language learning. This contributes to existing literature by showing how emotional and cognitive factors affect metacognitive processes in AI-tutored learning contexts. Notably, it contrasts with earlier studies that suggest negative emotions enhance metacognitive accuracy

(Massoni, 2014; Agadzhanyan & Castel, 2024), demonstrating that positive emotions can also support improved accuracy.

The article begins with an introduction that outlines the role of AI in education and the importance of feedback in learning. It then reviews studies on feedback and metacognition, highlighting a gap in understanding the effect of feedback timing within an AI-mediated environment. The Methods section describes the study design, participants, and the process of using conversational AI to deliver immediate and delayed feedback. The Results section presents findings on how feedback timing affects emotional states, cognitive load, and metacognitive accuracy. Finally, the Discussion interprets these findings, discusses theoretical and practical implications, and addresses limitations and future research directions.

1.1 Feedback and metacognition

Metacognition refers to the process of monitoring and regulating one's own performance (Kuhn & Dean, 2004). Feedback, defined as information provided by an agent (e.g., teacher, peer, book, parent, self, or experience) regarding aspects of a learner's performance or understanding (Hattie & Timperley, 2007), plays a critical role in this process. It informs learners about their performance, and it can significantly shape how learners evaluate their own learning – metacognitive judgment (Luo & Liu, 2023).

Nelson and Naren's Metacognitive Model (Nelson,1990) differentiate between two levels of cognitive processing: the object-level and the meta-level. The object-level refers to where the actual cognitive tasks happen, such as learning or reading. In the context of this study, this corresponds to participants completing the task of reading French sentences. While meta-level involves evaluating how well the task is being performed. In this study, this occurs when participants complete the metacognitive judgment questionnaire, where they assess their own performance. According to the model, information flows between these two levels in two directions. Monitoring is conceived as information flow from the object level to the meta-level, where participants evaluate their performance based on internal cues (self-assessment) and external feedback (whether immediate or delayed). This process forms the basis of their metacognitive judgments. On the other hand, information flowing from the meta-level to the object-level is called control and informs the object-level what to do next. Metacognitive judgments, which assess one's own learning and performance, fall under the monitoring aspect of this model. To be more specific, as they read, they assess their performance based on internal cues (self-assessment), and external feedback (immediate or delayed). This process forms their metacognitive judgment. Here are the key differences between metacognitive judgment and self-evaluation: In terms of timing, metacognitive judgment occurs during or immediately after specific parts of a task, whereas self-evaluation takes place after completing the entire task. In terms of focus, metacognitive judgment involves monitoring and predicting specific aspects of performance, while self-evaluation reflects on the overall quality of the performance.

Several studies have explored the relationship between feedback and metacognitive judgment, but their findings are inconsistent. For example, Haddara (2022) found that feedback had no effect on metacognitive sensitivity in a study on perceptual decision-making, even after seven days of training. In contrast, Luo and Liu (2023) found that trial-by-trial feedback enhanced metacognition in easy perceptual judgments but impaired it in difficult ones. Another study by Callender, Franco-Watkins, & Roberts (2016) showed that feedback in classroom settings helped students adjust their performance and judgments more effectively, leading to better calibration. Similarly, Geurten and Meulemans (2017) demonstrated that feedback improved the accuracy of children's metacognitive judgments by anchoring their predictions closer to their actual performance, particularly in memory tasks. Additionally, Urban & Urban (2021) finds that performance feedback improves accuracy in predictive judgments, helping children adjust their expectations after incorrect answers.

Building on this existing research, this study aims to further explore the role of feedback in metacognitive judgments, specifically looking into how the timing of feedback (immediate versus delayed) affects the accuracy of these judgments. Although the specific impact of feedback timing on metacognition has not been extensively studied, we believe that it could significantly shape how learners process feedback and adjust their self-evaluations. Immediate feedback may help learners to quickly adjust their understanding, helping them build more accurate self-assessment skills. On the other hand, delayed feedback gives learners time to reflect before receiving input, which can lead to more thoughtful self-evaluations. This is especially important in AI-based tutoring systems, where feedback timing can be carefully controlled to support and enhance the development of metacognitive skills.

We hypothesize that immediate feedback will lead to a more accurate metacognitive judgments compared to delayed feedback. This is because immediate corrections can help learners monitor their task performance in real time, leading to enhanced metacognitive calibration.

H1: Immediate feedback will result in more accurate metacognitive judgement than delayed feedback.

1.2 Emotion, cognitive load, and feedback

In a study, learners' affective states were significantly influenced by feedback from AutoTutor, an Intelligent Tutoring System (ITS) (Aghaei Pour et al., 2010). Specifically, positive feedback tends to induce "delight," while negative feedback often evokes "surprise." Meanwhile, in a study examining the effects of corrective feedback (CF) on learner affect in a computer-assisted language learning system, no significant difference in valence between a no-feedback group and a corrective feedback group was found (Bodnar et al., 2017). While these studies explore the influence of feedback type on emotional responses, they do not specifically address how the timing of feedback (immediate versus delayed) might affect valence. Since both feedback content and context can shape emotional outcomes, it is reasonable to assume that feedback timing could also play a crucial role. Therefore, we propose the following hypothesis: there will be a difference in valence between immediate and delayed feedback, as feedback timing may change learners' emotional responses. This leads to the following hypothesis:

H2a: There will be a difference in valence between immediate and delayed feedback.

Another study examines the relationship between feedback (specifically grades) and arousal, finding that the sign of feedback has a curvilinear (U-shaped) effect on arousal. This means that extreme feedback (either highly positive or highly negative) results in higher levels of arousal, whereas more moderate feedback results in lower arousal (Kluger, Lewinsohn, & Aiello, 1994). This suggests that the intensity or emotional impact of feedback plays a role in shaping learners' arousal levels. However, while the intensity of feedback has been explored, the timing of feedback-that is, whether feedback is presented immediately or after a delay-has received limited attention in terms of its impact on arousal levels. Since immediate and delayed feedback might differ in perceived intensity based on their timing, we propose the following hypothesis: there will be a difference in arousal between immediate and delayed feedback, as the timing may influence how learners emotionally process and respond to feedback. This leads to the following hypothesis:

H2b: There will be a difference in arousal between immediate and delayed feedback.

Research on the relationship between feedback and cognitive states shows that feedback may impact the way in which learners experience cognitive load. According to Cognitive Load Theory, cognitive resources are limited (Sweller, 2011), and feedback plays a role in how learners allocate these resources during learning. Explanatory feedback reduces extraneous cognitive load by providing detailed explanations that support learning. In contrast, corrective feedback increases extraneous load by offering only right-or-wrong judgments without additional guidance (Moreno, 2004). Strategy feedback tends to increase cognitive load by requiring more working memory resources to process complex problem-solving strategies, while outcome feedback imposes less cognitive load as it focuses only on the correctness of answers. The effectiveness of these feedback types, however, depends on the learner's cognitive capacity (Fyfe et al., 2015). Despite these insights, research on multimedia learning environments found no significant relationship between feedback type (simple or elaborate) and cognitive load (Lin et al., 2013). However, while these studies focus on the content and format of feedback, they do not consider how the timing of feedback—whether provided immediately or with a delay—might influence cognitive load. Since feedback timing can influence how learners process information, it may play a key role in how cognitive resources are allocated. Therefore, we propose that feedback timing (immediate versus delayed) may be associated with cognitive load, as it could impact how learners allocate and manage their cognitive resources during learning. Based on this, we present the following hypothesis:

H2c: There will be a difference in cognitive load between immediate and delayed feedback.

1.3 Emotion, cognitive load, and metacognition

The relationship between emotion and metacognition is still underexplored in the literature, and the findings that do exist tend to be context dependent. For example, Massoni (2014) observed that negative emotions may help metacognitive processes. However, the complexity of emotions, especially the variety of their components, may have differing effects on metacognition. This suggests that different types of emotion, even those with the same or oppoiste valence, may influence metacognition in different ways. Similarly, while Agadzhanyan and Castel (2024) demonstrated that negative emotional valence enhances both metacognitive judgments and memory performance, their study did not address how positive emotional valence might interact with metacognition, leaving a significant gap in understanding how different emotional charges could influence learning predictions. As a result, while there is enough evidence to suggest that a relationship exists, it remains difficult to propose a clear, consistent direction for this relationship. Expanding on this, Undorf, Söllner, and Bröder (2018) showed that emotional valence plays a significant role in metacognitive judgments. Their study showed that both positively and negatively charged words lead to higher judgment of learning compared to neutral words. This leads us to hypothesize that emotional valence, regardless of whether it is positive or negative, influences how individuals assess their own learning. Therefore, we propose the following hypothesis:

H3a: Valence is associated with the accuracy of metacognitive judgments.

Besides valence, another component of emotion is arousal. Garfinkel et al. (2013) suggested that heightened physiological arousal can negatively affect memory processes. They also proposed that it may also influence metacognitive judgments, such as confidence. According to their findings, the direction of this relationship depends on an individual's metacognitive insight and sensitivity to internal states. Thus, this leads us to hypothesize that arousal may influence metacognitive judgment, but the exact nature is very much context-dependent, thus requiring further investigation. Therefore, we propose the following hypothesis:

H3b: Arousal is associated with the accuracy of metacognitive judgments.

Another body of research has looked into the relationship between cognitive load and metacognitive judgments. The results of a meta-analysis synthesizing a number of studies on the relation between effort and monitoring judgments found a negative association between these two (Baars et al., 2020). However, that depends on the way learners invest their effort. In situations in which learners rely on data as a cue for data-driven regulation, higher cognitive load is often associated with lower confidence and more conservative metacognitive judgments. However, when learners use a goal-driven regulation strategy, allocating effort based on task importance or goals, the relationship between cognitive load and metacognitive judgments can change. Sometimes it becomes less negative or even positive. Such discrepancies in findings lead to the assumption that cognitive load influences the metacognitive judgment, though the specific direction of this influence depends on the regulation strategy used and the context of the learning task. This leads to the following hypothesis:

H3c: Cognitive load is associated with the accuracy of metacognitive judgments.

1.4 The Research Model

Figure 1 presents the research model based on the hypotheses shown above. The model shows the hypothesized relationship between feedback timing, emotional state (valence and arousal), cognitive state, and metacognitive judgment. It examines its direct impact on metacognitive judgment (H1), and its mediated effects through valence (H2a), arousal (H2b), and cognitive load (H2c), as well as the potential impact of these states on metacognitive judgment (H3a, H3b, H3c).

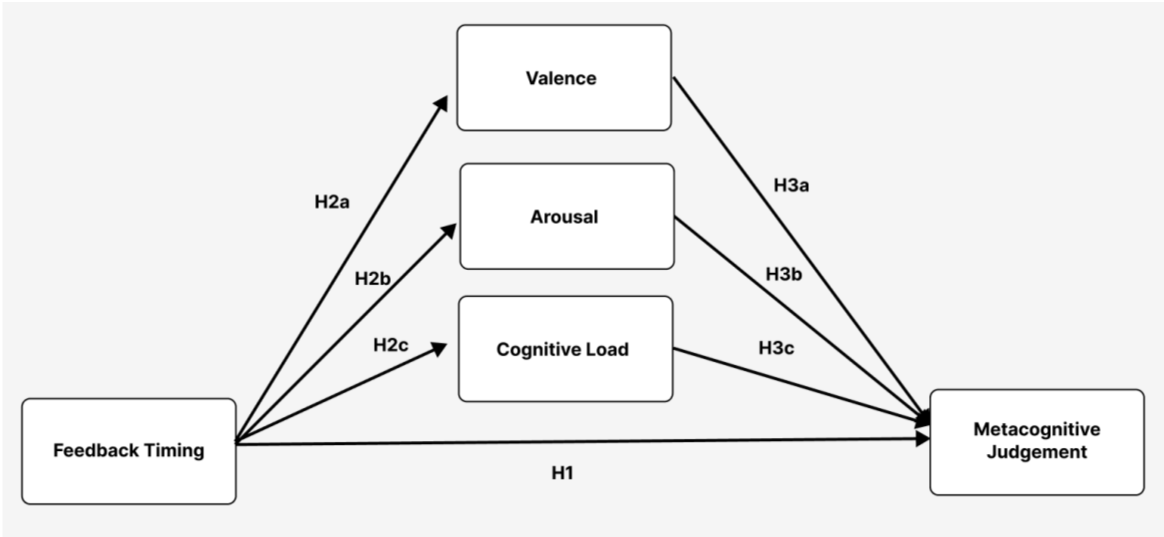

Figure 1. Research Model Illustrating the Relationships Between Feedback Timing, Emotional and Cognitive States, and Metacognitive Judgment

2. Methods

This study used a between-subjects design in a controlled laboratory setting to test the hypothesis regarding the impact of feedback timing (immediate versus delayed) on metacognitive judgment. Participants were randomly assigned to two conditions: the immediate group received feedback right after completing each sentence, while the delayed group received all corrective feedback at the end of the read-aloud task. Two mediators, emotional state (valence and arousal) and cognitive load, were also measured using a combination of self-reported scales and psychophysiological assessments to look into how these factors mediate the relationship between feedback timing and metacognitive judgment.

2.1 Participants

The participant demographics of this study include adult English speakers (18 years or older) living in Quebec, who were at the A2 level of French language proficiency according to the

Common European Framework of Reference for Languages (CEFR). This meant the participants could understand common sentences and expressions related to everyday needs, such as basic personal and family information, shopping, local geography, and employment. They should be able to complete simple tasks requiring straightforward information exchange on familiar topics and could describe basic aspects of their background, immediate environment, and needs. We utilized a convenience sampling method. Our inclusion criteria are that participants must not only live in Quebec and speak English. Additionally, they must be at the beginning of their journey in learning French, not necessarily as a second language but as an additional one. Those with intermediate or advanced proficiency in French were excluded to focus on beginner learners. Our sample size was 30 participants.

Participants were recruited both online and in-person through language institutions, cafes, and friends to reach the targeted population. Those showing interest in the experiment were screened for the beginner's criterion using a CEFR-based French language self-assessment. Each participant was offered a $30 honorarium to thank them for their contribution to the research.

2.2 Stimuli

In this study, we utilized French sentences as experimental stimuli, which were created using Tobii Pro Lab. The sentences were generated with the assistance of ChatGPT to ensure consistency and relevance across all three trials. For a comprehensive overview of the stimuli, including the specific texts and the prompts used to generate them, please refer to Appendix. Figures 2 and 3 show screenshots of what the stimuli look like, while Figure 4 illustrates what participants in the delayed condition would see.

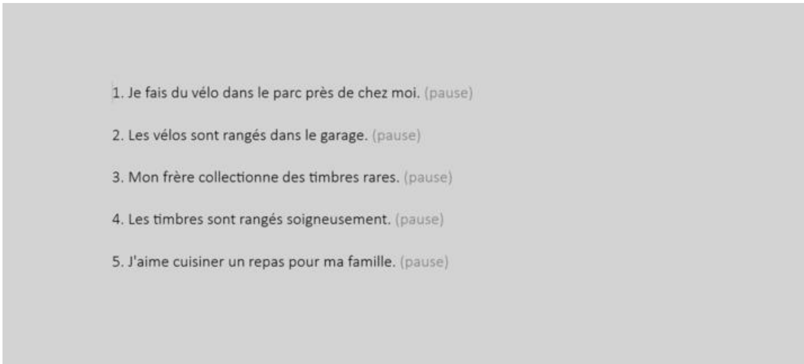

Figure 2. Screenshot of Experimental Stimuli (5 French sentences)

Figure 3. Screenshot of Experimental Stimuli (another 5 French sentences)

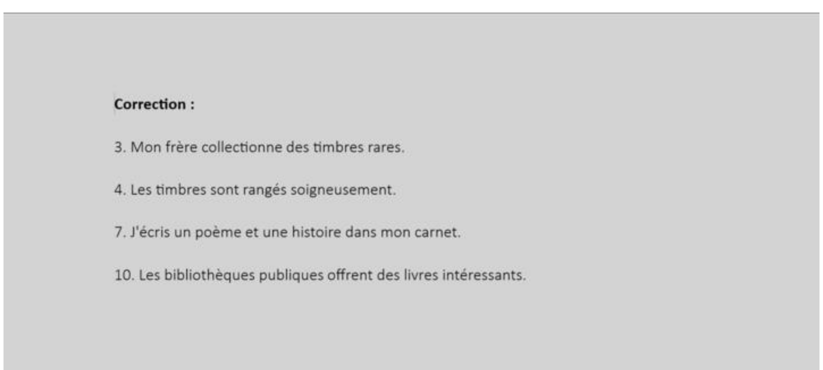

Figure 4. Correction Page Displayed to Participants in the Delayed Feedback Condition

The experimental manipulation focuses on the timing of feedback, not the content. All participants received the same feedback, where they were prompted to correct specific words, even if they hadn't actually made a mistake. In total, participants were asked to correct four specific words, which were distributed across different sentences. In the immediate feedback condition, participants received feedback on the targeted words right after reading the corresponding sentence. The AI tutor provided the correct pronunciation, and participants were then asked to repeat the whole sentence with the corrected pronunciation before moving on to the next one. In contrast, participants in the delayed feedback condition received feedback on the same four words only after completing all 10 sentences. Although the feedback timing is different between two conditions, the content of the feedback was the same for all participants.

The AI tool used in this experiment was OpenAI's GPT-4 model from ChatGPT, configured using prompt engineering. This method involves creating structured prompts to guide the AI's behavior, allowing the model to deliver adapted feedback on French pronunciation. We utilized the voice function of GPT-4 to enable spoken interactions. Details of the final prompt configurations used for generating the desired feedback can be found in the Appendix.

2.3 Procedure

When participants arrived at the lab, they were greeted by the researcher, who provided an overview of the study. Each participant then read and signed the consent form, so they knew their rights and what the study was for. Once consent was obtained, participants were randomly assigned to one of the experimental conditions (immediate or delayed). Participants then received detailed instructions about the tasks they needed to perform, and any questions they asked were explained to ensure clarity. The participants were seated in front of the Tobii Pro Lab (Tobii AB, Danderyd, Sweden) setup, where eye-tracking calibration was performed to ensure accurate data collection. After calibration, a baseline measurement was taken, which involved counting white squares. Following this, participants were provided with demographic information, including details about the language they learned, their educational level, and their employment status. The participants also need to complete a short warm-up exercise where they had to read 3 lines of French sentences, as appeared on the computer screen.

For the first task, participants in the immediate feedback condition read 10 French sentences displayed across two pages (see Figures 2 and 3), with 5 sentences per page. This feedback was provided immediately after participants finished reading each sentence that contained one of the four target words. The AI tutor would then offer the correct pronunciation of the specific word, asking participants to listen and repeat the pronunciation accurately before moving on to the next sentence. In contrast, participants in the delayed feedback condition read the same 10 sentences without receiving immediate corrections. After completing all the sentences, they were provided with a correction page that displayed only the sentences where pronunciation errors had been detected (see Figure 4). The AI tutor provided the correct pronunciation, and participants were then asked to repeat the entire sentence, incorporating the corrected pronunciation before moving on to the next one.

After completing the read task, participants were asked to fill out a brief questionnaire that was designed to measure their metacognitive judgments and emotions related to the task. The metacognitive judgment questionnaire focused on two aspects of pronunciation: comprehensibility (how easily their speech could be understood) and accentedness (how closely their pronunciation aligned with the native accent). We focused on pronunciation (comprehensibility and accentedness) because it is particularly challenging for learners to self-assess (Trofimovich et al., 2016). Unlike skills like listening or reading (Li & Zhang, 2021), learners often struggle to judge how closely their speech matches the target accent (accentedness) or how easy it is for others to understand them (comprehensibility). The self-assessment of metacognitive judgment on pronunciation was conducted using a slider scale labeled for comprehensibility (0 = hard to understand, 100 = easy to understand) and accentedness (0 = not accented at all, 100 = heavily accented) (Tsunemoto, 2022).

The procedure was followed a total of three times, to allow participants to perform multiple trials. Each trial involves a different set of 10 sentences so that there is sufficient data to analyze and to reduce variability in task performance. We used ChatGPT to generate all the texts that the participants read. The specific texts for each trial and how they were created can be found in Appendix: Stimuli Creation.

Finally, participants were thanked for their time and participation, and the compensation form was given out, and then they were escorted out of the lab.

2.4 Measures

Perceived cognitive load was assessed using a self-reported slider questionnaire based on the NASA TLX (Hart & Staveland, 1988), including questions such as: 1) 'How mentally demanding was the task?' and 2) 'How hard did you have to work to accomplish your level of performance?' Psychophysiological cognitive load was measured using pupillometry (Krejtz et al., 2018).

Perceived valence and arousal were measured with affective sliders (Betella & Verschure, 2016), while psychophysiological valence was assessed using FaceReader (Skiendziel, Rösch, & Schultheiss, 2019). Psychophysiological arousal was captured through average phasic electrodermal activity using the Cobalt Bluebox (Courtemanche et al., 2022).

Retrospective and prospective metacognitive judgments were calculated based on the difference between participants' self-reported performance on comprehensibility and accentedness and their actual performance, which was rated by three French speakers using the same scale. The average of the three raters' scores was taken as the actual performance since the inter-rater reliability ICC = 0.88. The inter-rater reliability was assessed using intraclass correlation.

Table 1 provides each construct measured in this study and their operationalization, including self-reported questionnaires, psychophysiological measures, and performance ratings.

Table 1. Construct and their operationalization.

Construct name	**Measure type**	**Description**	**Source Reference**
Metacognitive Judgment (Prospective & Retrospective)	Self-reported	a pre (prospective)-and post-task(retrospective) multi-item Likert-scale questionnaire Item 1: “Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.” Item 2: “Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.”	Tsunemoto, A (2022)
Perceived Cognitive Load	Self-reported	1. How mentally demanding was the task? 2. How hard did you have to work to accomplish your level of performance? Percentage slider: 0 = low, 100 = high *Selected from NASA TLX	Hart, S. G., & Staveland, L. E. (1988).

Table 1. Construct and their operationalization.

Cognitive Load	Psychophysio	Pupillometry	Krejtz et al. (2018)
Perceived Valence	Self-reported	Affective slider (1-item ranging from 0-100) "Move the slider to rate your level of valence felt during the read-aloud activity. Valence refers to the positive or negative character of the emotions you felt."	Betella, A., & Verschure, P. F. (2016).
Valence	Psychophysio	Valence inference from Facereader (Noldus)	Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019)
Perceived Arousal	Self-reported	Affective slider (1-item ranging from 0-100) "Move the slider to rate your level of arousal felt during the read-aloud activity. Arousal refers to the intensity of the emotions you felt."	Betella, A., & Verschure, P. F. (2016).
Arousal	Psychophysio	Average phasic electrodermal activity (Cobalt Bluebox)	Courtemanche et al (2022)

2.4 Data analysis

All statistical analyses were performed using SAS (version 3.81). The dataset included continuous variables for arousal, valence, and cognitive load (measured both through self-reported and psychophysiological measures), as well as prospective and retrospective metacognitive judgment, with accentedness and comprehensibility as the two evaluated aspects.

To test H1, a logistic regression was conducted. This method was used because the dependent variable, metacognitive judgment, was not normally distributed even after using log transformation. Therefore, a median split was performed. The value for median split is 16 (on a scale ranging from 0 to 55) for comprehensibility, and 14 (on a scale ranging from 0 to 72) for accentedness.

To test H3a, a logistic regression was conducted. Again, this method was used because the dependent variable, metacognitive judgment, was not normally distributed even after using log transformation. Thus, a median split was performed. The value for median split is 16 (on a scale ranging from 0 to 55) for comprehensibility, and 14 (on a scale ranging from 0 to 72) for accentedness. The same approach was applied to test H3b, as well as H3c.

To test if there is a difference in valence between immediate feedback and delayed feedback, (H2a) a linear regression with random intercepts was conducted, as the dependent variable, valence, was normally distributed. The same approach was used to test the hypotheses that differences exist in arousal and cognitive load between the immediate and delayed feedback conditions (H2b & H2c).

Statistical significance was determined at the $\alpha = 0.05$ level, with all p-values, estimates, and standard errors reported.

3. Results

3.1 Descriptive statistics

Table 2 presents the descriptive statistics for both the immediate and delayed feedback conditions across various measures. It provides a summary of how participants' emotional states, cognitive load, and self-assessment of their pronunciation varied between the immediate and delayed feedback conditions.

Table 2. Descriptive statistics				
Variable	Immediate feedback		Delayed feedback	
	M	SD	M	SD
Perceived Arousal	61.55	14.11	64.54	21.22
Physiological Arousal	-4.63	4.12	-3.92	3.45
Perceived Valence	63.37	17.42	67.00	19.42
Physiological Valence	-.062	.15	-.104	.117
Perceived Cognitive Load	31.59	19.36	40.01	21.53
Physiological Cognitive Load	2.89	.28	2.88	.26
Prospective Metacognitive Judgment (comprehensibility)	18.18	16.29	19.40	15.26
Prospective Metacognitive Judgment (accentedness)	22.64	20.17	18.45	15.71
Retrospective Metacognitive Judgment (comprehensibility)	20.15	15.12	19.75	14.99
Retrospective Metacognitive Judgment (accentedness).	22.57	20.41	17.31	16.94

3.2 Hypothesis testing

H1 suggests that immediate feedback will result in more accurate metacognitive judgement than delayed feedback. Since no significant relationship was found between feedback timing and retrospective metacognitive judgment, we will focus only on the results of prospective metacognitive judgment. The logistic regression analyses, as shown in Table 3, shows that there is no significant effect of feedback timing on the accuracy in metacognitive judgment for either comprehensibility (p = .663) or accentedness (p = .459). Therefore, H1 is not supported.

Table 3.
Logistic Regression: Impact of Feedback Timing on Perspective Metacognitive Judgment

DV	estimate	SE	P
Metacognitive Accuracy (comprehensibility)	-.3109	.737	.675
Metacognitive Accuracy (accentedness)	.089	.868	.919

Note: Modeling the probability of having higher metacognitive accuracy

H2a suggests that there will be a difference in valence between immediate and delayed feedback. As shown in Table 4, the linear regression results show that there was no significant difference in perceived valence (Perceived Valence, p = .537) or arousal (Perceived Arousal, p = .579) when comparing immediate feedback condition to delayed feedback conditions. In addition, the analysis of cognitive load, as measured by TLX (Perceived Cognitive Load, p = .263) and pupillometry (Physiological Cognitive Load, p = .813), show no significant differences between feedback conditions. Therefore, H2a, H2b, and H2c are not supported. The findings, summarized in Table 4, indicate that feedback timing did not significantly affect valence, arousal, or cognitive load.

Table 4.
Linear regression: Impact of Feedback Timing on Valence, Arousal, and Cognitive Load

DV	Estimate	SE	p value
Perceived Arousal	-3.33	5.97	.579
Physiological Arousal	-.719	1.49	.631
Perceived Valence	-3.91	6.28	.537
Physiological Valence	.029	.047	.547
Perceived Cognitive Load	-7.81	6.89	.263
Physiological Cognitive Load	.023	.098	.813

Note: Negative estimates show that the dependent variable is lower in the immediate feedback condition compared to the delayed feedback condition. Positive estimates show that the dependent variable is higher in the immediate feedback condition compared to the delayed feedback condition. Comparisons are based on two-tailed tests.

Again, since no significant relationship was found between valence, arousal, cognitive load, and retrospective metacognitive judgment, we will focus only on the results of prospective metacognitive judgment. To further explore the effects of these factors on metacognitive judgment, we conducted logistic regression analyses. As shown in Table 5, the results partially support the hypothesis about the relationship between these factors and metacognitive judgment. A significant positive association was found between perceived valence and accuracy in metacognitive judgment for comprehensibility ($\beta = .047$, SE = .02, $p = .024$). While the association for accentedness was not significant ($\beta = .04$, SE = .022, $p = .07$), the p-value suggests a weak trend towards significance.

H3a suggests that valence is associated with the accuracy of metacognitive judgments. In line with the hypothesis, the results show that higher perceived valence is associated with greater accuracy in metacognitive judgments. As shown in Table 5, the model estimates the probability that metacognitive accuracy (comprehensibility) equals 0. Since the lower value of metacognitive accuracy (comprehensibility) indicates higher accuracy, a positive estimate (β =.047) suggests that higher perceived valence is linked to improved accuracy for comprehensibility. However, no significant relationships were found between perceived or physiological arousal and metacognitive accuracy for either comprehensibility or accentedness. Therefore, H3b is not supported and H3a is partially supported.

H3c suggests that cognitive load is associated with the accuracy of metacognitive judgments. Physiological Cognitive Load showed a significant effect on prospective metacognitive judgment accuracy for accentedness ($\beta = 3.791$, SE = 1.81, $p = .042$). However, no significant effect was found for comprehensibility ($p = .865$). Therefore, H3c is partially supported. Specifically, higher cognitive load being associated with greater accuracy in metacognitive judgments about accentedness.

Table 5.
Logistic Regression: Impact of Valence, Arousal, and Cognitive Load on Perspective Metacognitive Judgment

	Metacognitive accuracy (comprehensibility)			Metacognitive accuracy (Accentedness)		
IV	estimate	SE	P	estimate	SE	P
Perceived Arousal	.023	.018	.205	.017	.021	.403
Physiological Arousal	-.086	.094	.365	1.08	.115	.353
Perceived Valence	.047	.02	.024	.04	.022	.07
Physiological Valence	.999	2.97	.738	-.894	3.41	.794
Perceived Cognitive Load	.021	.017	.215	-.019	.019	.338
Physiological Cognitive Load	-1.62	1.45	.269	3.791	1.81	.042

Note: Modeling the probability having higher metacognitive accuracy

Table 6 presents a summary of hypothesis testing results for the study. The table shows each hypothesis and the extent to which the results supported the hypothesis. Feedback timing did not influence the metacognitive judgment (H1), valence (H2a), arousal (H2b), and cognitive load (H2c). However, valence (H3a) and cognitive load (H3c) partially supported the hypothesis that they influence metacognitive judgment accuracy.

Table 6. hypothesis Testing

Hypothesis	From	To	estimate	p-Value	Status
H1	Feedback Timing	Metacognitive Judgment - Comprehensibility	-.3109	0.663	not support
H1	Feedback Timing	Metacognitive Judgment - Accentedness	.089	0.459	not support
H2a	Feedback Timing	Perceived Valence	-3.91	.537	not support
H2a	Feedback Timing	Valence	.029	.547	not support
H2b	Feedback Timing	Perceived Arousal	-3.33	.579	not support
H2b	Feedback Timing	Arousal	-.719	.631	not support
H2c	Feedback Timing	Perceived Cognitive Load	-7.81	.263	not support
H2c	Feedback Timing	Cognitive Load	.023	0.813	not support
H3a	Perceived Valence	Metacognitive Judgment - Comprehensibility	.047	.024	supported
H3a	Perceived Valence	Metacognitive Judgment - Accentedness	.04	.07	not support
H3a	Valence	Metacognitive Judgment - Comprehensibility	0.999	.738	not support
H3a	Valence	Metacognitive Judgment - Accentedness	-.894	.794	not support
H3b	Perceived Arousal	Metacognitive Judgment - Comprehensibility	.023	.205	not support
H3b	Perceived Arousal	Metacognitive Judgment - Accentedness	.017	.403	not support
H3b	Arousal	Metacognitive Judgment - Comprehensibility	-.086	.365	not support
H3b	Arousal	Metacognitive Judgment - Accentedness	1.08	.353	not support
H3c	Perceived Cognitive Load	Metacognitive Judgment - Comprehensibility	.021	.215	not support
H3c	Perceived Cognitive Load	Metacognitive Judgment - Accentedness	-.019	.338	not support
H3c	Cognitive Load	Metacognitive Judgment - Comprehensibility	-1.62	.269	not support
H3c	Cognitive Load	Metacognitive Judgment - Accentedness	3.791	.042	supported

Figure 5 presents the Validated Research Model, which illustrates the relationships tested in the study and shows which paths were fully supported, partially supported, or not supported by the results.

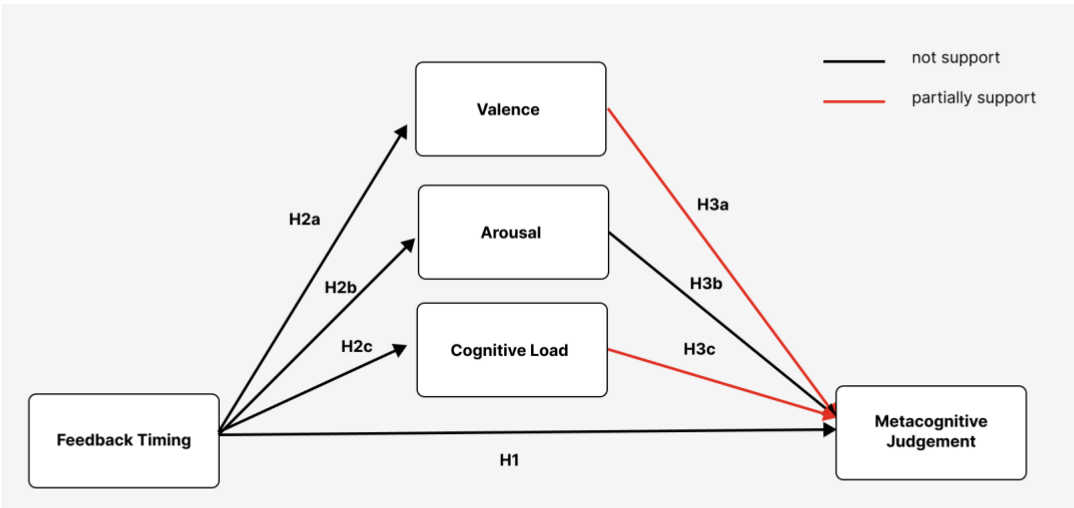

Figure 5. Validated Research Model Illustrating the Relationships Between Feedback Timing, Emotional and Cognitive States, and Metacognitive Judgment

4. Discussion

The main findings showed that feedback timing (immediate versus delayed) did not significantly impact learners' emotional states (valence and arousal), cognitive load, or the accuracy of metacognitive judgments for comprehensibility and accentedness in language learning. However, an interesting result was that higher perceived emotional valence was associated with higher metacognitive judgment accuracy, suggesting that learners who experienced higher positive emotions tended to evaluate their pronunciation performance more accurately. Additionally, cognitive load showed a significant relationship with metacognitive accuracy for accentedness, implying that higher cognitive effort might enhance learners' awareness of their pronunciation. These results highlight that while the timing of feedback might not directly influence metacognitive judgments, emotional states and cognitive load are important factors in learners' metacognitive evaluation processes.

Contribution to Theory

This study makes two key theoretical contributions. First, it introduces a novel perspective on feedback timing (immediate versus delayed) in AI-mediated language learning. It extends our understanding of how conversational AI affects metacognitive processes like self-monitoring and self-evaluation in pronunciation. While research has examined the role of feedback in traditional educational settings (Fu & Li, 2022; Rassaei, 2023; Corral et al., 2021), this study focuses on the timing of feedback in AI-based learning and its influence on metacognition, addressing a critical gap in the existing literature.

Second, this study highlights how emotional states and cognitive load impact metacognitive judgment accuracy, showing that higher emotional valence and increased cognitive load led to more accurate metacognitive judgment. In the context of AI-mediated learning, this finding is important because it shows that emotional and cognitive state lead to a more accurate self-assessment of French pronunciation. Interestingly, these results contradict with some existing research suggesting that negative emotions enhance metacognitive accuracy (Massoni, 2014; Agadzhanyan & Castel, 2024). Instead, our study found that positive emotion can enhance metacognitive accuracy. Similarly, higher cognitive load and better metacognitive accuracy contracts theories that lower cognitive effort can improve self-assessment (Baars et al., 2020).

However, the finding suggests that it is context-specific, which has been shown in prior research.

Contributions to Methods

The study's methodological contribution is the innovative use of the latest ChatGPT-4 model as AI tutor for second language learning. Unlike traditional educational tools, using ChatGPT-4 model as AI tutor allows for adapted and interactive feedback, delivered in both immediate and delayed conditions. It shows how AI can be adapted to meet specific educational goals. This innovative approach shows the potential of advanced AI tools to enhance both language learning and metacognitive research.

Furthermore, it offers a replicable framework for future studies about AI-mediated learning. This framework includes clear guidance for integrating AI tools like ChatGPT into experimental design so that researchers can replicate the study to explore similar questions in diverse learning contexts.

Contributions to Practice

AI tutor could be useful for students of any age trying to learn a second language. It offers a structured and interactive way to practice pronunciation, based on each learner's needs and proficiency level. Immediate feedback promotes active correction by addressing errors on the spot. On the other hand, delayed feedback encourages self-reflection and helps learners to critically evaluate their overall performance and recognize patterns in their mistakes. This combination of real-time corrective and self-assessment makes the AI tutor especially helpful for beginner learners.

Beyond the use of Gen AI as language learning tool, the AI tools also show potential in healthcare, such as aiding speech therapists in helping patients regain speaking abilities. The flexibility of the feedback options allows patients to practice at their own pace. Therefore, they would feel more at ease with practicing their speaking skills and have time to reflect on their progress. This creates a comfortable and effective learning environment, fostering continuous improvement and helping patients in their recovery journey.

In addition to its role in assisting therapists in helping patients regain speaking abilities, the AI tutor also has potential in other fields. For instance, in primary education, it can be a valuable tool for elementary school students, especially those struggling with reading or pronunciation. By providing personalized and adaptive feedback, AI tutors can help bridge the gap in early literacy and language development, making young learners more confident in their speaking skills and eventually achieve fluency.

Limitation and Future Research

This study can be further explored in two main areas. Firstly, the relatively short trials duration might have limited the difference observed between immediate and delayed conditions. Since participants only read 10 French sentence per trial in both condition, which may have not provided enough opportunities for feedback to fully impact the immediate group. This might have potential minimized the variation between two feedback groups. Future studies could extend the trials' duration so that a more noticeable difference could be potentially observed. Secondly, the sample size of 30 participants, constrained by time and funding, may have influenced the statistical power of the results. In future research, recruiting a larger sample would enhance the reliability and significance of the findings. By addressing the shorter trial duration and small sample size, future studies can better capture potential differences and lead to more insights.

Conclusion

This study aimed to explore the impact of feedback timing on metacognitive judgment in AI-mediated language learning. It addresses the gap in understanding how immediate and delayed feedback influence emotional states, cognitive load, and self-assessment in pronunciation. Using conversational AI, the study showed that while feedback timing itself may not significantly affect metacognitive accuracy, emotional valence, and cognitive load have impacts on improving learners' self-assessment. In the future, AI in language learning could focus on emotional and cognitive engagement to increase metacognitive accuracy. By creating more personalized feedback that takes into account learners' emotional and cognitive state, AI learning tools can better support self-regulated learning and improve the overall learning experience.

References

Statista. (2023). *Generative Artificial Intelligence (AI)*. Retrieved October 3, 2024, from https://www.statista.com/study/133112/generative-artificial-intelligence-ai/

Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty years of personalized language learning. Educational Technology & Society, 24(1), 205-222.

Taj, I. H., Ali, F., Sipra, M. A., & Ahmad, W. (2017). Effect of technology enhanced language learning on vocabulary acquisition of EFL learners. International Journal of Applied Linguistics and English Literature, 6(3), 262-272.

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. International handbook of metacognition and learning technologies, 1-16.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay–retention effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(1), 80.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112.

Webb, J. M., Diana, E. M., Luft, P., Brooks, E. W., & Brennan, E. L. (1997). Influence of pedagogical expertise and feedback on assessing student comprehension from nonverbal behavior. The Journal of Educational Research, 91(2), 89-97.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. Review of educational research, 58(1), 79-97.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Memory & cognition, 36(3), 604-616

Rassaei, E. (2023). The interplay between corrective feedback timing and foreign language anxiety in L2 development. Language Teaching Research, 13621688231195141.

Corral, D., Carpenter, S. K., & Clingan-Siverly, S. (2021). The effects of immediate versus delayed feedback on complex concept learning. Quarterly Journal of Experimental Psychology, 74(4), 786-799.

Kuhn, D., & Dean, Jr, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. Theory into practice, 43(4), 268-273.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 918.

Tsunemoto, A., Trofimovich, P., Blanchet, J., Bertrand, J., & Kennedy, S. (2022). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. Foreign Language Annals, 55(1), 135-154.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology, 52, 139-183.

Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. Studies in Second Language Acquisition, 44(1), 2-34.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. PloS one, 13(9), e0203629.

Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. PloS one, 14(10), e0223905.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. PloS one, 11(2), e0148037.

Bluebox: Courtemanche, F., Sénécal, S., Fredette, M., & Léger, P.-M. (2022). In H. Montréal (Ed.), COBALT-Bluebox: Multimodal user data wireless synchronization and acquisition system.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. PloS one, 11(2), e0148037.

Massoni, S. (2014). Emotion as a boost to metacognition: How worry enhances the quality of confidence. Consciousness and cognition, 29, 189-198.

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. Metacognition and learning, 11, 215-235.

Luo, T., & Liu, C. (2023). The impact of feedback on metacognition: Enhancing in easy tasks, impeding in difficult ones. Consciousness and Cognition, 116, 103601.

Haddara, N., & Rahnev, D. (2022). The impact of feedback on perceptual decision-making and metacognition: Reduction in bias but no change in sensitivity. Psychological Science, 33(2), 259-275.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). Handbook of metacognition in education.

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In Psychology of learning and motivation (Vol. 26, pp. 125-173). Academic Press.

Brown, A. L. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In Weinert, F. E. & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 65–116). Lawrence Erlbaum Associates.

Baars, M., & Wijnia, L. (2020). The Relation Between Student's Effort and Monitoring Judgments During Learning: A Meta-analysis. Educational Psychology Review. Retrieved from http://hdl.handle.net/1765/133563

Garfinkel, S. N., Barrett, A. B., Minati, L., Dolan, R. J., Seth, A. K., & Critchley, H. D. (2013). What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity. Psychophysiology, 50(6), 505-512.

Geurten, M., & Meulemans, T. (2016). The effect of feedback on children's metacognitive judgments: a heuristic account. Journal of Cognitive Psychology, 29(2), 184–201. https://doi.org/10.1080/20445911.2016.1229669

Urban, K., & Urban, M. (2021). Anchoring Effect of Performance Feedback on Accuracy of Metacognitive Monitoring in Preschool Children. Europe's Journal of Psychology, 17(1), 104-118. https://doi.org/10.5964/ejop.2397

Agadzhanyan, K., & Castel, A. D. (2024). The effect of emotional valence and font size on metacognition and memory. Memory, 32(2), 252–263. https://doi.org/10.1080/09658211.2024.2307919

Undorf, M., Söllner, A. & Bröder, A. Simultaneous utilization of multiple cues in judgments of learning. Mem Cogn 46, 507–519 (2018). https://doi.org/10.3758/s13421-017-0780-6

Sweller, J. (2011). Cognitive load theory.

Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter?. Computers & Education, 67, 239-249.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. Instructional science, 32(1), 99-113.

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: the importance of working memory capacity. Instructional Science, 43, 73-91.

Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017). Learner affect in computerised L2 oral grammar practice with corrective feedback. Computer Assisted Language Learning, 30(3–4), 223–246. https://doi.org/10.1080/09588221.2017.1302964

Aghaei Pour, P., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. In V. Aleven, J. Kay, & J. Mostow (Eds.), Intelligent tutoring systems (Lecture Notes in Computer Science, Vol. 6094). Springer. https://doi.org/10.1007/978-3-642-13388-6_31

Kluger, A. N., Lewinsohn, S., & Aiello, J. R. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. Organizational Behavior and Human Decision Processes, 60(2), 276-299.

Tobii AB (2024). Tobii Pro Lab [Computer software]. Danderyd, Sweden: Tobii AB.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self-and other-perception of second language speech. Bilingualism: Language and Cognition, 19(1), 122-140.

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. Language Testing, 38(2), 189-218.

Footnotes

This article is in preparation for Journal Consciousness and Cognition, and it has been formatted accordingly.

Chapter 3

GEN AI: Revolutionizing Language Learning and Beyond

As of August 2024, Open AI has confirmed that ChatGPT now has more than 200 million active weekly users (VentureBeat, 2024). With the release of ChatGPT-4's advanced voice functions, OpenAI continues to push the boundaries of what's possible with AI (OpenAI, 2023). Among the distinctive features of Chat GPT-4 is its ability to speak multiple languages, therefore making the model possible to be used in language learning (Tulsiani, 2024). Based on this, in our study we developed a French language tutor that uses the capabilities of ChatGPT to give users individual feedback with the aim of improving their language skills. The article also discusses how such AI-based tutors can be used in the context of different domain-specific training applications for enhancing communication skills and supporting specialist training.

Using ChatGPT-4's advanced language capabilities, we created a French tutor designed specifically to help users practice their pronunciation. Two versions of the AI tutor were created: one that provides immediate feedback, correcting mispronounced words after each sentence, while the other one delivers feedback after the learner finish reading a set of sentences (e.g., after reading 10 sentences). The development process was well-structured and iterative. We started by creating basic prompts for pronunciation correction and then conducted internal tests to test its accuracy. We then let the real user test the tools and gather feedback about their experience. Observing users' interactions with the AI tutor helped us identify areas for improvement, which we addressed through multiple iterations. These updates enhance the tools clarity, speed, and precision to better support beginner French learners. Participants highlighted the tutor's accuracy and clarity, often think that feedback was relevant, well-explained and helpful. This shows AI tutor's potential as an accessible and effective tool for language learning.

AI tutor could be useful for students of any age trying to learn a second language. It offers a structured and interactive way to practice pronunciation, based on each learners' needs and proficiency level. Immediate feedback promotes active correction by addressing errors on the spot. On the other hand, delayed feedback encourages self-reflection and helps learners to critically evaluate their overall performance and recognize patterns in their mistakes. This combination of real-time corrective and self-assessment makes the AI tutor especially helpful for beginner learners.

The AI tutor also shows great potential in the field of healthcare, particularly in supporting speech therapists in helping patients regain their speaking abilities (Price, Lubniewski, & Du, 2024). One of its unique strengths is that it provides both immediate and delayed feedback, which means patients have the flexibility of practicing their speaking skills using either one of the feedback options at their own time. Therefore, they would feel more at ease with practicing their speaking skills and have time to reflect on their progress. This creates a comfortable and effective learning environment, fostering continuous improvement and helping patients in their recovery journey.

In addition to its role in assisting therapists in helping patients regain speaking abilities, the AI tutor also has potential in other fields. For instance, in primary education, it can be a valuable tool for elementary school students, especially those struggling with reading or pronunciation. By providing personalized and adaptive feedback, AI tutors can help bridge the gap in early literacy and language development, making young learners more confident in their speaking skills and eventually achieve fluency.

One of the advantages of the AI tutor is that it was developed in-house on the OpenAI platform, making it cost-effective and accessible to different types of organization (ranging from small educational institutions to large enterprises). For example, small educational institutions can integrate the AI tutor into their language program. Since it offers personalized practice without adding to instructors' workloads (Miyazoe, 2024). While for larger organizations, the AI tutor can act as a role-play partner helping employees to refine their communication skills (Ajjan, 2024). For instance, it could provide a real-life scenario, such as acting as a client or business partner during a training session. So that employees can practice conversation in a safe environment.

AI technologies, like ChatGPT-4, have created new opportunities for personalized learning through its advanced voice function and multilingual support. This article examined the application of such AI tutors across different sectors to offer customize feedback and support specialized training. Since AI tutors offer both immediate and delayed feedback, it allows users to not only correct errors in real-time but also reflect on their progress. This flexible feature supports both learning and self-assessment. It is particularly beneficial in the health and education sectors, supporting speech therapy and language programs. Moreover, the in-house development of the AI tutors makes them accessible to various organizations, providing a low-cost, flexible solution for different training needs.

References

Franzen, C. (2024, August 29). OpenAI says ChatGPT now has 200M users. VentureBeat. https://venturebeat.com/ai/openai-says-chatgpt-now-has-200m-users/

OpenAI. (2023, September 25). ChatGPT can now see, hear, and speak. OpenAI. Retrieved December 3, 2024, from https://openai.com/index/chatgpt-can-now-see-hear-and-speak/#OpenAI

Tulsiani, R. (2024, January 12). Transforming language learning with AI. eLearning Industry. Retrieved December 3, 2024, from https://elearningindustry.com/transforming-language-learning-with-ai

Price, L., Lubniewski, K., & Du, Y. (2024, March 10). Using ChatGPT to create treatment materials. The ASHA Leader. Retrieved December 3, 2024, from https://leader.pubs.asha.org/do/10.1044/leader.EOAI.29032024.chatgbt-therapy-materials-slp.26/full/

Ajjan, H. (2024, January 25). What ChatGPT's voice and image capabilities mean for educators. Harvard Business Publishing Education. Retrieved December 3, 2024, from https://hbsp.harvard.edu/inspiring-minds/what-chatgpts-voice-and-image-capabilities-mean-for-educators

Miyazoe, T. (2024, May 6). Practical insights: Incorporating ChatGPT in language education and beyond. EDUCAUSE Review. Retrieved December 3, 2024, from https://er.educause.edu/articles/2024/5/practical-insights-incorporating-chatgpt-in-language-education-and-beyond

Chapter 4

Conclusion

As AI continues to transform industries, its impact on education is growing significantly. AI-driven technologies have the potential to create flexible learning experiences, pushing the boundaries of traditional teaching methods. The way these AI systems provide feedback is crucial in helping learners to assess their own performance. Despite the importance of feedback in learning, little is known about how the timing of feedback (immediate versus delayed) affects learners' ability to evaluate their own progress (metacognitive judgment). This study examined the impact of feedback timing (immediate versus delayed) on metacognitive judgments within an AI-mediated system. Two key research questions were addressed: first, to what extent does feedback timing influences metacognitive judgment accuracy, and second, to what extent learners' emotional and cognitive states (valence, arousal, and cognitive load) impact the relationship between feedback timing and metacognitive judgment. These questions aimed to fill the gap in literature, by investigating the role of feedback timing in AI-driven learning contexts.

The study's findings showed that feedback timing (immediate versus delayed) did not significantly impact learners' emotional states (valence and arousal), cognitive load, or the accuracy of metacognitive judgments related to the comprehensibility and accentedness of their French pronunciation. However, higher emotional valence was associated with more accurate metacognitive judgments for comprehensibility, and higher cognitive load was associated with more accurate metacognitive judgments for accentedness. These results suggest that while feedback timing itself may not directly influence metacognitive accuracy, emotional and cognitive factors play a critical role in shaping learners' self-assessments during AI-mediated language learning.

The findings from this study provide both theoretical and methodological contributions, as well as practical implications. Theoretically, the study expands the understanding of metacognition by exploring how AI-mediated feedback influences self-monitoring and self-evaluation processes. Specifically, it fills the gap the in literature by showing that while feedback timing might not directly impact metacognitive judgment, factors like emotional valence and arousal play an important role in learners' ability to self-assess their performance. (Agadzhanyan & Castel, 2024; Baars et al., 2020). This suggests that the influence of feedback timing maybe be more complex than previously assumed, it might require considering factors like emotional and cognitive state, instead of only considering when the feedback is delivered. The study also contributes to the field of educational AI by showing how advanced AI tools can enhance language learning and support more accurate metacognitive judgment, which opens the door to future studies in AI in educational contexts.

From a practical perspective, the results suggest that AI based educational tools should prioritize cognitive and emotional engagement, rather than focusing on feedback timing as a key factor to increase learners' self-assessment. Methodologically, this study demonstrates the effective use of GPT-4 as a conversational AI tutor, providing adapted feedback in both real-time and delayed formats. This approach offers a replicable framework for future research in AI-mediated learning environments.

Future research should address the limitations of this study to gain deeper insights. Firstly, extending the trial duration or increasing the amount of feedback in the immediate condition could better demonstrate the differences between immediate and delayed feedback. In addition, increasing the sample size would increase the statistical power of the results, allowing for a stronger conclusion. While feedback in this study was not personalized, future research could focus on customizing feedback based on individual characteristics, such as emotional state, cognitive load or performance. These studies could look into whether personalized feedback leads to better immediate learning outcomes or long-term metacognitive development. Future studies could also explore feedback timing and personalization in other learning contexts, such as problem solving or critical thinking. Exploring how feedback interacts with different cognitive tasks could help us understand metacognition across different domains. Lastly, future research could explore the long-term effect of feedback timing on metacognition. By applying longitudinal method, we could track learners over time to see how immediate versus delayed feedback could impact long term self-assessment. These studies could show whether learners benefit from a specific type of feedback over time or if they adapt their strategies differently according to the type of feedback they receive.

In conclusion, this research enhances our understanding of how feedback timing, emotional states, and cognitive load impact metacognitive judgments in AI-mediated language learning. While the study revealed that feedback timing did not significantly impact learners' emotional responses, cognitive load, or metacognitive judgments of French pronunciation, it highlights the importance of emotional and cognitive engagement in enhancing learners' metacognitive accuracy. The findings suggest that factors such as emotional valence and cognitive load are more influential in shaping learners' ability to monitor and evaluate their performance (metacognitive accuracy) than previously assumed, instead of focusing on feedback timing alone. As AI continues to evolve in the educational landscape, these insights will be essential for developing more effective, responsive learning tools.

Bibliography

Statista. (2023). *Generative Artificial Intelligence (AI)*. Retrieved October 3, 2024, from https://www.statista.com/study/133112/generative-artificial-intelligence-ai/

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In Psychology of learning and motivation (Vol. 26, pp. 125-173). Academic Press.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. Psychonomic Bulletin & Review, 14, 225-229.

Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. Studies in Second Language Acquisition, 44(1), 2-34.

Rassaei, E. (2023). The interplay between corrective feedback timing and foreign language anxiety in L2 development. Language Teaching Research, 13621688231195141.

Corral, D., Carpenter, S. K., & Clingan-Siverly, S. (2021). The effects of immediate versus delayed feedback on complex concept learning. Quarterly Journal of Experimental Psychology, 74(4), 786-799.

Saltzman, I. J. (1951). Delay of reward and human verbal learning. Journal of experimental psychology, 41(6), 437.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. Review of educational research, 58(1), 79-97.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Memory & cognition, 36(3), 604-616.

Luo, T., & Liu, C. (2023). The impact of feedback on metacognition: Enhancing in easy tasks, impeding in difficult ones. Consciousness and Cognition, 116, 103601.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 918.

Aghaei Pour, P., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. In V. Aleven, J. Kay, & J. Mostow (Eds.), Intelligent tutoring systems (Lecture Notes in Computer Science, Vol. 6094). Springer. https://doi.org/10.1007/978-3-642-13388-6_31

Kluger, A. N., Lewinsohn, S., & Aiello, J. R. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. Organizational Behavior and Human Decision Processes, 60(2), 276-299.

Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty years of personalized language learning. Educational Technology & Society, 24(1), 205-222.

Taj, I. H., Ali, F., Sipra, M. A., & Ahmad, W. (2017). Effect of technology enhanced language learning on vocabulary acquisition of EFL learners. International Journal of Applied Linguistics and English Literature, 6(3), 262-272.

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. International handbook of metacognition and learning technologies, 1-16.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay–retention effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(1), 80.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112.

Webb, J. M., Diana, E. M., Luft, P., Brooks, E. W., & Brennan, E. L. (1997). Influence of pedagogical expertise and feedback on assessing student comprehension from nonverbal behavior. The Journal of Educational Research, 91(2), 89-97.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Memory & cognition, 36(3), 604-616

Kuhn, D., & Dean, Jr, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. Theory into practice, 43(4), 268-273.

Tsunemoto, A., Trofimovich, P., Blanchet, J., Bertrand, J., & Kennedy, S. (2022). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. Foreign Language Annals, 55(1), 135-154.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology, 52, 139-183.

Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. Studies in Second Language Acquisition, 44(1), 2-34.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. PloS one, 13(9), e0203629.

Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. PloS one, 14(10), e0223905.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. PloS one, 11(2), e0148037.

Bluebox: Courtemanche, F., Sénécal, S., Fredette, M., & Léger, P.-M. (2022). In H. Montréal (Ed.), COBALT-Bluebox: Multimodal user data wireless synchronization and acquisition system.

Massoni, S. (2014). Emotion as a boost to metacognition: How worry enhances the quality of confidence. Consciousness and cognition, 29, 189-198.

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. Metacognition and learning, 11, 215-235.

Luo, T., & Liu, C. (2023). The impact of feedback on metacognition: Enhancing in easy tasks, impeding in difficult ones. Consciousness and Cognition, 116, 103601.

Haddara, N., & Rahnev, D. (2022). The impact of feedback on perceptual decision-making and metacognition: Reduction in bias but no change in sensitivity. Psychological Science, 33(2), 259-275.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). Handbook of metacognition in education.

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In Psychology of learning and motivation (Vol. 26, pp. 125-173). Academic Press.

Brown, A. L. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In Weinert, F. E. & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 65–116). Lawrence Erlbaum Associates.

Baars, M., & Wijnia, L. (2020). The Relation Between Student's Effort and Monitoring Judgments During Learning: A Meta-analysis. Educational Psychology Review. Retrieved from http://hdl.handle.net/1765/133563

Garfinkel, S. N., Barrett, A. B., Minati, L., Dolan, R. J., Seth, A. K., & Critchley, H. D. (2013). What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity. Psychophysiology, 50(6), 505-512.

Geurten, M., & Meulemans, T. (2016). The effect of feedback on children's metacognitive judgments: a heuristic account. Journal of Cognitive Psychology, 29(2), 184–201. https://doi.org/10.1080/20445911.2016.1229669

Urban, K., & Urban, M. (2021). Anchoring Effect of Performance Feedback on Accuracy of Metacognitive Monitoring in Preschool Children. Europe's Journal of Psychology, 17(1), 104-118. https://doi.org/10.5964/ejop.2397

Agadzhanyan, K., & Castel, A. D. (2024). The effect of emotional valence and font size on metacognition and memory. Memory, 32(2), 252–263. https://doi.org/10.1080/09658211.2024.2307919

Undorf, M., Söllner, A. & Bröder, A. Simultaneous utilization of multiple cues in judgments of learning. Mem Cogn 46, 507–519 (2018). https://doi.org/10.3758/s13421-017-0780-6

Sweller, J. (2011). Cognitive load theory.

Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter?. Computers & Education, 67, 239-249.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. Instructional science, 32(1), 99-113.

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: the importance of working memory capacity. Instructional Science, 43, 73-91.

Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017). Learner affect in computerised L2 oral grammar practice with corrective feedback. Computer Assisted Language Learning, 30(3–4), 223–246. https://doi.org/10.1080/09588221.2017.1302964

Aghaei Pour, P., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. In V. Aleven, J. Kay, & J. Mostow (Eds.), Intelligent tutoring systems (Lecture Notes in Computer Science, Vol. 6094). Springer. https://doi.org/10.1007/978-3-642-13388-6_31

Kluger, A. N., Lewinsohn, S., & Aiello, J. R. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. Organizational Behavior and Human Decision Processes, 60(2), 276-299.

Tobii AB (2024). Tobii Pro Lab [Computer software]. Danderyd, Sweden: Tobii AB.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self-and other-perception of second language speech. Bilingualism: Language and Cognition, 19(1), 122-140.

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. Language Testing, 38(2), 189-218.

Franzen, C. (2024, August 29). OpenAI says ChatGPT now has 200M users. VentureBeat. https://venturebeat.com/ai/openai-says-chatgpt-now-has-200m-users/

Appendices

Request for authorization to submit in the form of articles

Office of the Registrar

3000 chemin de la Côte-Sainte-Catherine Montreal, Quebec, Canada H3T 2A7

HEC MONTRÉAL

1. Student

Asikaer Nadila — 11334393

Last name, First name — HEC ID number

2. Teaching department responsible : Information Technology

Master of Science — User Experience

Program of study — Specialisation

3. List of proposed articles

Author(s) : Nadila Asikaer, Sylvain Senecal, Constantinos K. Coursaris

Title : AI-Driven Feedback Timing and Its Impact on Learners' Metacognitive Judgments

Journal or book:

Current status of article: ☐ published ☐ submitted for publication ☑ in preparation

Author(s) :

Title :

Journal or book:

Current status of article: ☐ published ☐ submitted for publication ☐ in preparation

Author(s) :

Title :

Journal or book:

Current status of article: ☐ published ☐ submitted for publication ☐ in preparation

Author(s) :

Title :

Journal or book:

Current status of article: ☐ published ☐ submitted for publication ☐ in preparation

4. Student's signature

Nadila Asikaer — NadilaAsikaer — Oct 23,2024

Student's name — Signature — Date

5. Approval by Research Director/Co-Directors

Constantinos Coursaris — Constantinos Coursaris — Signature numérique de Constantinos Coursaris Date : 2024.10.24 14:46:23 -04'00'

Director's name — Signature — Date

Sylvain Senecal — Sylvain Senecal — Signature numérique de Sylvain Senecal Date : 2024.10.23 12:50:58 -04'00'

Co-Director's name (if applicable) — Signature — Date

6. Decision and signature of Program Director

Decision: ☐ Accepted ☐ Refused

Program Director — Signature — Date

Figure 1. Request for authorization to submit in the form of articles

Table 1. Data Dictionary

Accronyms	Description	Comments
T1, T2, T3	trial order	3 trilas are counterbalanced
P0	pre task	P0: before task, P1: phase1, P2: phase2
P1	post task	
TLX-MED_1	Perceived cognitive load: item 1	mental demand
TLX-EFF_1	Perceived cognitive load: item 2	effort
P0_MetaC_1	prospective metacognitive judgment:item1	comprehensibility
P0_MetaC_2	prospective metacognitive judgment:item2	accentedness
P1_MetaC_1	retrospective metacognitive judgment :item1	comprehensibility
P1_MetaC_2	retrospective metacognitive judgment :item2	accentedness
AS_arousal	Perceived arousal	
AS_valence	perceived valence	
PriorExp_VA-1	Prior experience with Voice Assistants item 1	
PriorExp_VA-2	Prior experience with Voice Assistants item 2	
PriorExp_VA-3	Prior experience with Voice Assistants item 3	
Lang	language	
Edu	educational level	
Employ	employment status	

Table 2. Data Dictionary

Source	Variable name	Label
Photobooth	phasic_gsr_	EDA
Photobooth	k_dispersion_	K Coefficient
Photobooth	LeftSize_Pupil_	Pupil size left
Photobooth	RightSize_Pupil_	Pupil size right
Photobooth	Valence	Valence from Face Reader
Photobooth	GTE	Gaze transition Entropy
Tobii	Fixation duration	Fixation duration
Tobii	Fixation count	Fixation count
Tobii	Saccade amplitude	Saccade amplitude
Tobii	Saccade frequency	Saccade frequency
Tobii	Time to first fixation	Time to first fixation
tobii	task performance: efficiency	Task completion time
Noted by the RA (LRDG sheet)	task performance: effectiveness	Task success ratio

Table 3. Data Dictionary

Variable name	Label
StartDate	Date de commencement
EndDate	Date de fin
Status	Type de réponse
IPAddress	Adresse IP
Progress	Progrès
Duration (in seconds)	Durée (en secondes)
Finished	Terminé
RecordedDate	Date enregistrée
ResponseId	Identifiant de la réponse
RecipientLastName	Nom du destinataire
RecipientFirstName	Prénom du destinataire
RecipientEmail	Adresse e-mail du destinataire
ExternalReference	Référence externe de données
LocationLatitude	Latitude de l'emplacement
LocationLongitude	Longitude de l'emplacement
DistributionChannel	Canal de distribution
UserLanguage	Langue de l'utilisateur

Table 3. Data Dictionary

CONDITION	Enter the condition that is provided by the moderator.
PX	Enter the participant number that is provided by the moderator.
LHQ_Lang_1	Indicate your native language(s) and any other languages you have studied or learned. - Language 1:
LHQ_Lang_11	Indicate your native language(s) and any other languages you have studied or learned. - Language 2:
LHQ_Lang_12	Indicate your native language(s) and any other languages you have studied or learned. - Language 3:
LHQ_Lang_13	Indicate your native language(s) and any other languages you have studied or learned. - Language 4:
LHQ_Lang_14	Indicate your native language(s) and any other languages you have studied or learned. - Language 5:
LHQ_Lang_15	Indicate your native language(s) and any other languages you have studied or learned. - Language 6:
LHQ_Lang_16	Indicate your native language(s) and any other languages you have studied or learned. - Language 7:
LHQ_Lang_17	Indicate your native language(s) and any other languages you have studied or learned. - Language 8:
LHQ_Lang_18	Indicate your native language(s) and any other languages you have studied or learned. - Language 9:
LHQ_Lang_19	Indicate your native language(s) and any other languages you have studied or learned. - Language 10:
Edu	Indicate the highest level of education you have completed.
Employ	Indicate your current occupation.
PriorExp_VA-1	Have you ever used any voice assistant before (e.g., Siri, ChatGPT, Google Assistant, Amazon Alexa, Microsoft Cortana) ?
PriorExp_VA-2	How frequently do you interact with voice assistants in your daily life ?
PriorExp_VA-3	How comfortable are you with interacting with voice assistants compared to other forms of technology interfaces (e.g., websites, mobile apps) ?

Table 3. Data Dictionary

T1_P0_MetaC_1_1	Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.
T1_P0_MetaC_2_1	Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.
T1_Activity	Enter the activity number that is provided by the moderator.
T1_P1_MetaC_1_1	Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.
T1_P1_MetaC_2_1	Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.
T1_P1_P-Feedback_1	Rate your level of agreement or disagreement with the following statements : - I would consider the feedback justified.
T1_P1_P-Feedback_2	Rate your level of agreement or disagreement with the following statements : - I would consider the feedback useful.
T1_P1_TLX-MED_1	How mentally demanding was the read-aloud activity ?
T1_P1_TLX-EFF_1	How hard did you have to work to accomplish your level of performance ?
T1_P1_AS-Arousal_1	Move the slider to rate your level of arousal felt during the read-aloud activity. Arousal refers to the intensity of the emotions you felt.
T1_P1_AS-Valence_1	Move the slider to rate your level of valence felt during the read-aloud activity. Valence refers to the positive or negative character of the emotions you felt.
T1_P2_AS-Arousal_1	Move the slider to rate your level of arousal felt during the learning activity. Arousal refers to the intensity of the emotions you felt.
T1_P2_AS-Valence_1	Move the slider to rate your level of valence felt during the learning activity. Valence refers to the positive or negative character of the emotions you felt.

Table 3. Data Dictionary

T2_P0_MetaC_1_1	Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.
T2_P0_MetaC_2_1	Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.
T2_Activity	Enter the activity number that is provided by the moderator.
T2_P1_MetaC_1_1	Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.
T2_P1_MetaC_2_1	Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.
T2_P1_P-Feedback_1	Rate your level of agreement or disagreement with the following statements : - I would consider the feedback justified.
T2_P1_P-Feedback_2	Rate your level of agreement or disagreement with the following statements : - I would consider the feedback useful.
T2_P1_TLX-MED_1	How mentally demanding was the read-aloud activity ?
T2_P1_TLX-EFF_1	How hard did you have to work to accomplish your level of performance ?
T2_P1_AS-Arousal_1	Move the slider to rate your level of arousal felt during the read-aloud activity. Arousal refers to the intensity of the emotions you felt.
T2_P1_AS-Valence_1	Move the slider to rate your level of valence felt during the read-aloud activity. Valence refers to the positive or negative character of the emotions you felt.
T2_P2_AS-Arousal_1	Move the slider to rate your level of arousal felt during the learning activity. Arousal refers to the intensity of the emotions you felt.
T2_P2_AS-Valence_1	Move the slider to rate your level of valence felt during the learning activity. Valence refers to the positive or negative character of the emotions you felt.

Table 3. Data Dictionary

T3_P0_MetaC_1_1	Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.
T3_P0_MetaC_2_1	Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.
T3_Activity	Enter the activity number that is provided by the moderator.
T3_P1_MetaC_1_1	Move the slider to rate the level of comprehensibility. Comprehensibility refers to how easily someone else can understand you speaking French.
T3_P1_MetaC_2_1	Move the slider to rate your level of accentedness. Accentedness refers to the degree to which a speaker's French pronunciation deviates from the native or standard pronunciation.
T3_P1_P-Feedback_1	Rate your level of agreement or disagreement with the following statements : - I would consider the feedback justified.
T3_P1_P-Feedback_2	Rate your level of agreement or disagreement with the following statements : - I would consider the feedback useful.
T3_P1_TLX-MED_1	How mentally demanding was the read-aloud activity ?
T3_P1_TLX-EFF_1	How hard did you have to work to accomplish your level of performance ?
T3_P1_AS-Arousal_1	Move the slider to rate your level of arousal felt during the read-aloud activity. Arousal refers to the intensity of the emotions you felt.
T3_P1_AS-Valence_1	Move the slider to rate your level of valence felt during the read-aloud activity. Valence refers to the positive or negative character of the emotions you felt.
T3_P2_AS-Arousal_1	Move the slider to rate your level of arousal felt during the learning activity. Arousal refers to the intensity of the emotions you felt.
T3_P2_AS-Valence_1	Move the slider to rate your level of valence felt during the learning activity. Valence refers to the positive or negative character of the emotions you felt.

Table 3. Data Dictionary

POST_Usefulness-1_1	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor is pretty much what I need to learn French.
POST_Usefulness-1_2	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor adequately meets my information needs.
POST_Usefulness-1_3	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The information provided by the AI Tutor is effective.
POST_Usefulness-2_1	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor allows me to interact with it to receive tailored information.
POST_Usefulness-2_2	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor has interactive features, which help me learn French.
POST_Usefulness-2_3	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - I can interact with the AI Tutor in order to get information tailored to my specific needs.
POST_Usefulness-3_1	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - When I use the AI Tutor, there is very little waiting time between my actions and the response.
POST_Usefulness-3_2	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor loads quickly.
POST_Usefulness-3_3	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor takes long to load.
POST_Relevance_1	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor is relevant for my language learning course.
POST_Relevance_2	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor would help me progress in my French pronunciation.
POST_Relevance_3	Rate your level of agreement or disagreement with the following statements concerning your overall experience with the AI Tutor : - The AI Tutor allows me to receive personalized learning that I cannot find in another place.
POST_ReUse_NPS_GROUP	If the need arose, what is the probability that you would use an AI Tutor for learning French pronunciation ? - Group
POST_ReUse	If the need arose, what is the probability that you would use an AI Tutor for learning French pronunciation ?

Stimuli Creation

We used ChatGPT to generate all the text that the participants will read. Below are the specific texts for each trial and an explanation of how they were created.

Prompt 1. (masculine/feminine)

Instruction: Please create a 15-sentence paragraph in French (the theme is introducing yourself and your family). Each pair of sentences should contain one word that appears in both masculine and feminine forms. For example, in sentence one: "un étudiant," and in sentence two: "une étudiante." Number each sentence and ensure that each sentence has exactly 10 words. The text should be easy to understand for A1-level French learners.

Sentence

1. Bonjour, je m'appelle Marie et je suis **étudiante**.
2. Mon frère Pierre est **étudiant**.
3. J'ai une sœur qui est **gentille**.
4. Mon père est un homme **gentil** et **intelligent**.
5. Ma mère, une femme **intelligente**, travaille dans une école.
6. Ma tante médecin est très **occupée** .
7. J'ai un cousin qui est **sportif**.
8. Ma cousine **sportive** joue au tennis.
9. Mon ami est un garçon **sérieux** qui travaille beaucoup.
10. Mon amie, une fille **sérieuse**, aime étudier.

Prompt 2. (singular/plural)

Instruction: Please create a 15-sentence paragraph in French (the theme is about hobby). Each pair of sentences should contain one word that appears in singular form in one sentence, and plural forms in another sentence. Number each sentence and ensure that each sentence has exactly 10 words. The text should be easy to understand for A1-level French learners.

Sentence:

1. Je fais du **vélo** dans le parc près de chez moi.
2. Les **vélos** sont rangés dans le garage.
3. Mon frère collectionne des **timbres** rares.
4. Les **timbres** sont rangés soigneusement.
5. J'aime cuisiner un **repas** pour ma famille.
6. Les **repas** du dimanche sont toujours spéciaux.
7. J'écris un **poème** et une histoire dans mon carnet.
8. Les **poèmes** sont inspirés par la nature.
9. J'aime lire des livres de ma **bibliothèque**.
10. Les **bibliothèques** publiques offrent des livres intéressants.

Prompt 3. (Contractions)

Instruction: Please create a 15-sentence paragraph in French (the theme is where I live) where each sentence contains the word "au", "à la", "à l'" or "aux". Number each sentence. Each sentence must contain exactly 10 words. The text should be easy to understand for A1-level French learners.

Sentence:

1. J'habite dans un petit village **au** bord de la mer.
2. Il y a un grand parc **au** centre de la ville.
3. Les enfants jouent souvent **aux** jeux dans ce parc.
4. Je vais **au** marché chaque samedi.
5. **À l'**école, les élèves apprennent le français et les mathématiques.
6. Nous dînons souvent **au** restaurant près de la maison.
7. Les voisins se retrouvent **au** café pour discuter.
8. Le dimanche, nous allons **à la** plage.
9. Il y a une bibliothèque **au** coin de ma rue.
10. J'adore me promener **au** parc avec mon chien.

Prompt 4. (warm up - number)

Instruction: Please create a 5-sentence paragraph in French (the theme is the address) where all the sentences must contain numbers written in French. Number each sentence. Each sentence must contain exactly 10 words. The text should be easy to understand for A1-level French learners.

Sentence:

1. J'habite au numéro **vingt** de la rue principale.
2. Ma voisine vit au numéro **trente-deux**, juste à côté.
3. Mon immeuble compte **huit** étages.