International Metaverse Association

Metaverse Convergence Research Institute

ISSN 3059-0442

Journal of Virtual Convergence Research

Volume 2 Number 2 Apr. 2026

Received: 28 January 2026. Accepted: 22 April 2026

Journal of Virtual Convergence Research
2026, VOL. 2, NO. 2, 23~97.
https://intlmetaverse.com/home/journal?lang=en

International Metaverse Association &
Metaverse Convergence Research Institute

A Study on the Interaction Effects of Avatar-Based Online Lectures: Focusing on Real-Time Q&A, Response Speed, and Language Style

TaeHwan Lee

Graduate School of Metaverse, Sogang University

luisfynn1@sogang.ac.kr

Abstract

Online learning often faces limitations due to unidirectional communication, where unresolved inquiries can disrupt learning flow and lead to student attrition. To address this, this study conducted a 2×2 factorial experiment ($N = 52$) utilizing an AI avatar-based real-time Q&A system, manipulating response speed (fast vs. slow) and linguistic style (conversational vs. formal). The findings indicated that while the real-time Q&A function did not significantly improve mean learning achievement, it demonstrated a "stabilization effect" by reducing the variance in achievement scores. Although main effects for the two variables were largely non-significant, a crossover interaction pattern with a medium effect size was observed for social presence and cognitive load, with social presence being notably higher in the conversational-fast and formal-slow conditions. Despite its statistical limitations as a small-scale pilot study, this research serves as significant preliminary evidence for the potential effectiveness of AI avatar-based interaction design in enhancing the online educational experience.

Keywords : AI Avatar, Real-Time Interaction, Response Latency, Linguistic Style, Learning Outcomes, Flow, Social Presence, Cognitive Load Theory

A Study on the Interaction Effects of Avatar-Based Online Lectures: Focusing on Real-Time Q&A, Response Speed, and Language Style

1. Introduction

The onset of the COVID-19 pandemic between 2020 and 2023 served as a critical turning point, establishing online lectures as a ubiquitous form of education. Furthermore, since the emergence of ChatGPT in late 2022, the rapid advancement of artificial intelligence (AI) has further accelerated this evolution toward personalized and adaptive learning environments. Despite the flexibility of online education, it continues to grapple with the inherent limitations of unidirectional communication. This one-way nature poses a significant risk: if learners fail to grasp specific concepts or terms in real-time, their comprehension of subsequent content is often compromised.

When learners are forced to pause a lecture to seek external information, the resulting search time and verification efforts lead to a "disruption of flow" and an increase in extraneous cognitive load. These factors ultimately contribute to diminished motivation and higher student attrition rates. To address these challenges, this study proposes an AI-driven avatar-based system for real-time Q&A during online lectures. The primary objective is to maintain learning immersion and mitigate cognitive load by providing immediate clarification, while simultaneously enhancing social presence through avatar interaction to improve overall learning outcomes.

Furthermore, this research moves beyond mere statistical significance to quantitatively examine the effect sizes of interaction design variables. By exploring the practical magnitude and direction of these impacts, this study aims to provide an empirical foundation for optimizing variable combinations and sample size estimations in future large-scale applications of AI-based pedagogical agents.

2. Theoretical Background

2.1 The Evolution of Online Learning and the Emergence of AI Avatars

Over the past several decades, online education has established itself as a mainstream pedagogical approach within higher education and vocational training. Recently, this field has entered a transformative era of innovation, driven by the rapid advancement of artificial intelligence (AI) and the strategic integration of AI avatars (Islam & Wang, 2025). These technological breakthroughs are currently spearheading the expansion of the global AI market, including the educational sector, which is projected to maintain a compound annual growth rate (CAGR) of over 30% (Islam & Wang, 2025).

AI avatars are increasingly gaining traction across various digital platforms, revolutionizing traditional content production workflows. In the media industry, this technology enables the creation of educational lectures without the physical presence of an instructor, the production of advertisements without hiring human talent, and the broadcasting of news without on-site anchors. A notable example of this trend is the UK-based platform Synthesia, which allows users to generate high-quality videos simply by inputting text. The platform utilizes sophisticated voice synthesis and lip-syncing technologies to animate chosen avatars. Such solutions are now being rapidly adopted in the field of educational content development, offering a scalable and efficient alternative to traditional video production methods.

2.2 Limitations of Online-Based Learning

The most significant advantage of online learning lies in its temporal and spatial flexibility. It is no longer an uncommon sight to observe individuals engaging with lectures via smartphones in public spaces, such as subways or buses. Utilizing "fragmented time"—including commutes and lunch breaks—for educational purposes has become a pervasive part of daily life.

However, despite the accessibility and flexibility offered by online platforms, a structural limitation persists: most instructional systems focus predominantly on the unidirectional transmission of information. This one-way delivery can exacerbate learner isolation and significantly diminish engagement (Briz-Ponce et al., 2017). Crucially, when questions arising during the learning process are not addressed immediately, the resulting "disruption of flow" serves as a primary driver of reduced learning efficiency and increased attrition rates (Briz-Ponce et al., 2017). Consequently, the development of interactive systems utilizing AI avatars is expected to be a decisive factor in mitigating these limitations and maximizing overall learning effectiveness.

2.3 Research Necessity and Core Research Gap

Recent literature has substantiated the potential value of avatars in enhancing the quality of online learning, reporting significant improvements in learner attention and positive affect (Xia et al., 2024). Specifically, Kodani et al. (2025) demonstrated that the embodied anthropomorphism of pedagogical agents fosters superior educational outcomes. However, existing studies have predominantly focused on the visual attributes of avatars or generalized social presence, leaving a critical theoretical lacuna regarding the micro-level dimensions that constitute interaction quality. In particular, the integrated effects of Response Latency and Linguistic Style (Tone)—two pivotal factors shaping the dynamics of interaction—on learners' cognitive-affective states and terminal learning outcomes remain under-explored (Ukenova et al., 2025).

To bridge this gap, the current study empirically analyzes how the temporal and linguistic dimensions of interaction influence the learner experience within a real-time Q&A framework. In this research, response speed (fast vs. slow) and linguistic style (conversational vs. formal) are manipulated as independent variables. Response speed is operationalized as the latency between query recognition and output generation, while linguistic style is systematically varied through sentence endings and expressions while maintaining semantic consistency across conditions.

Furthermore, this study posits that these two variables influence learning outcomes not only through direct pathways but also through indirect pathways mediated by three psychological constructs: Flow, Social Presence, and Cognitive Load. By adopting this structural approach, the research moves beyond merely identifying optimal conditions; it seeks to elucidate the internal mechanisms through which interaction design variables reconfigure the learner experience and pedagogical efficacy.

3. Review of Prior Research

This study constructs an integrated theoretical framework by synthesizing three core theories to elucidate the mechanisms linking the independent variables—Response Speed, Linguistic Style, and the presence of Q&A—to the mediating and dependent variables.

Flow Theory (Csikszentmihalyi, 1991): Flow theory provides the theoretical basis for the pathway between Response Speed and learning outcomes. It posits that immediate feedback is a critical prerequisite for maintaining and intensifying learning immersion (flow). In this study, the promptness of the AI avatar's response is hypothesized to sustain the learner's cognitive momentum, thereby preventing the "disruption of flow" and enhancing academic performance.

Cognitive Load Theory (Sweller, 1988) and the Personalization Principle (Mayer, 2021): These theories substantiate the impact of Linguistic Style on extraneous cognitive load. According to Mayer's personalization principle, information presented in a conversational rather than a formal tone reduces the cognitive effort required for processing, thereby optimizing the allocation of working memory resources. This study examines how linguistic variations modulate cognitive load to indirectly influence learning efficacy.

Social Presence Theory (Biocca et al., 2003): This theory explains how interaction modalities dictate the degree to which a learner perceives an AI avatar as a "social entity." Social presence serves as a pivotal psychological bridge, where a higher sense of co-presence and psychological connection with the avatar leads to increased engagement and improved achievement.

Synthesis of the Integrated Model: The structural configuration of this research is defined by the specific roles of each variable within this integrated framework. Response Speed primarily operates through the Flow and Social Presence pathways, while Linguistic Style is predominantly linked to Cognitive Load and Social Presence. The availability of a Q&A system is posited to influence terminal learning outcomes through all three mediating constructs—Flow, Cognitive Load, and Social Presence—as illustrated in Figure 2.

3.1 Flow Theory and Immediacy of Interaction

Flow theory, established by Csikszentmihalyi (1991), describes a psychological state of optimal experience in which an individual becomes completely absorbed in an activity. A critical precursor to inducing a state of flow is the provision of clear goals and immediate, unambiguous feedback (Csikszentmihalyi, 1991; Sherry, 2004). Such feedback allows learners to maintain a delicate balance between the challenge of the task and their own skill level, facilitating total concentration on the current activity (Csikszentmihalyi, 1991). Furthermore, Sherry (2004) demonstrated that flow in media environments is intrinsically linked to enjoyment, suggesting that immersive experiences serve as a powerful catalyst for positive learning motivation.

To operationalize this theoretical requirement for immediacy within a real-time Q&A environment, this study implemented a system where AI-generated responses were segmented at the sentence level using the Python re package and subsequently processed through real-time Text-to-Speech (TTS) synthesis. This technical configuration was designed to minimize the perceived waiting time for the learner.

When response latency increases, learners lose the sense of interactional immediacy, which directly obstructs "concentration on the task"—a core component of the flow state (Csikszentmihalyi, 1991). Such delays disrupt the cognitive momentum and learning flow, leading to increased subjective temporal pressure and frustration (Sherry, 2004). The significance of this immersion is further supported by Georgiou and Kyza (2018), who empirically established a positive correlation between immersion levels and conceptual learning outcomes in location-based augmented reality (AR) environments. In summary, the immediate feedback emphasized by flow theory is realized through rapid response speeds, which are hypothesized to enhance both learner engagement and academic achievement.

3.2 Cognitive Load Theory (CLT) and the Personalization Principle

Cognitive Load Theory (CLT), as posited by Sweller (1988), asserts that human working memory capacity is inherently finite. To optimize learning efficiency, instructional design must minimize extraneous cognitive load—the unnecessary mental effort unrelated to the actual difficulty of the content (intrinsic load) (Sweller, 1988). Applying CLT to multimedia learning environments, Mayer (2021) formulated the Personalization Principle, which focuses on optimizing the linguistic delivery of instructional materials.

The Personalization Principle maintains that presenting information in a conversational style, rather than a formal tone, mitigates the learner’s cognitive load and enhances comprehension (Mayer, 2021). This effect occurs because a conversational approach encourages learners to perceive the AI avatar as a "social partner" rather than a mere information conduit. This social framing prompts deeper cognitive processing while simultaneously reducing the mental effort required for information decoding (Mayer, 2021).

Furthermore, Ukenova et al. (2025) underscored the importance of linguistic structure alignment and sentiment-driven expressions as critical strategies for improving avatar-based learning systems. Their findings suggest that the systematic manipulation of linguistic style exerts a direct influence on pedagogical efficacy (Ukenova et al., 2025). By fostering rapport and minimizing extraneous cognitive load, a conversational style is hypothesized to significantly improve overall learning efficiency within the AI avatar-mediated environment.

3.3 AI Avatars, Social Presence, and Learning Efficacy

Beyond mere visual representation, avatars serve as pivotal interfaces that instill a sense of social presence and catalyze interaction within digital learning environments. Xia et al. (2024) substantiated the potential of avatars to enhance learner attention and positive affect while mitigating social anxiety, thereby elevating the overall quality of online education. Similarly, Pang et al. (2023) explored the capacity of avatar-based platforms to facilitate social interaction in collaborative settings, such as virtual chemistry laboratory sessions.

The psychological impact of avatar design is further elucidated by Kodani et al. (2025), who investigated the pedagogical effects of android avatars. Their findings indicate that while positive perceptions of an avatar's anthropomorphism and competence significantly improve subjective educational outcomes, perceived discomfort exerts a deleterious effect. This underscores the critical importance of strategic avatar design in shaping learners' psychological responses. Furthermore, Herbert and Dołżycka (2024) emphasized the pedagogical roles of artificial educational agents and visual avatars, focusing on their influence on learning outcomes, perception, and satisfaction.

In the context of the metaverse, Islam and Wang (2025) reviewed how AI integration and customization options reconfigure the learning experience, while Sinlapanuntakul and Zachry (2025) addressed the specific effects of avatar representation in video-mediated collaborative interactions. Collectively, these studies suggest that the implementation strategy and interaction quality of avatars are fundamental to the construction of social presence. This perspective aligns with the framework proposed by Briz-Ponce et al. (2017), which asserts that such interactional nuances ultimately dictate learners' overall behavioral intentions and technology acceptance.

3.4 Interaction Mechanisms between Temporal Efficiency and Linguistic Rapport

The response speed and linguistic expressions integrated into this study's online lecture system are hypothesized to significantly influence both social presence and learning outcomes. While response speed serves as a metric for the seamlessness and fluidity of the interaction, linguistic style functions as a primary driver for establishing rapport and familiarity between the learner and the avatar.

When an avatar's response is delayed, learners are likely to perceive the system as unresponsive or technically deficient. Such latency often triggers redundant behaviors, such as refreshing the page or re-entering the same query while the system is still processing. These actions necessitate cognitive processing unrelated to the actual learning material, thereby imposing an extraneous cognitive load and causing a significant disruption of flow.

According to Kodani et al. (2025), these delays induce learner discomfort, which subsequently leads to a negative evaluation of the avatar's utility. Furthermore, technical dissatisfaction arising from latency may negate the positive psychological effects typically associated with a conversational style, thereby undermining both social presence and pedagogical efficacy.

Consequently, this study aims to examine the nuanced interplay between response speed and linguistic style. Given the complexity of human-computer interaction, these two factors may not always exert a consistent linear influence; rather, they may exhibit complex, interactive patterns in real-world data. Nevertheless, it is hypothesized that the most pronounced learning effects will occur when these two variables achieve an optimal harmony, facilitating a high-quality, social-cognitive experience for the learner.

4. Research Hypotheses

To investigate the pedagogical and psychological effects of AI avatar-based interactions, this study initially categorizes participants into two primary conditions based on the availability of a real-time Q&A system. Within the group receiving the Q&A

service, a 2×2 between-subjects factorial design was employed, manipulating two independent variables: response speed and linguistic style.This configuration resulted in a total of five experimental groups—one control group (no Q&A) and four experimental subgroups (Fast-Conversational, Fast-Formal, Slow-Conversational, and Slow-Formal). Based on this design, the following hypotheses were formulated to examine the impact of these interaction variables on learning outcomes and psychological experiences:

4.1 Basic Effects of Interaction and Mediating Hypotheses

H1: The group utilizing the real-time Q&A system will exhibit significantly higher learning performance than the control group.

According to Flow Theory, learners maintain a state of immersion when provided with immediate and unambiguous feedback during an activity, which leads to deeper information processing and enhanced learning outcomes (Csikszentmihalyi, 1991; Sherry, 2004). In the context of online learning, real-time Q&A functionality serves as a critical tool that prevents "disruption of flow" by allowing learners to resolve uncertainties instantaneously, thereby enabling sustained concentration on the lecture content. Furthermore, Cognitive Load Theory (CLT) posits that extraneous cognitive load, arising from unresolved questions or comprehension failures, is a primary barrier to learning efficiency (Sweller, 1988; Mayer, 2021). By facilitating real-time clarification, the Q&A system mitigates this unnecessary cognitive burden, allowing learners to allocate their cognitive resources solely to the instructional material. Consequently, it is hypothesized that the real-time Q&A group will demonstrate superior academic performance compared to the control group.

H2: The effect of real-time Q&A on learning performance will be positively mediated by flow and social presence, and negatively mediated by cognitive load.

Flow theory suggests that immediate feedback and task-focused concentration induce a state of immersion, which serves as a core mechanism for enhancing learning performance (Csikszentmihalyi, 1991). By utilizing real-time Q&A to obtain immediate answers to inquiries, learners can remain fully dedicated to the learning task, which is expected to exert a positive influence on their achievement.

Simultaneously, Social Presence Theory maintains that a higher perception of an interaction partner's social existence fosters greater learner engagement and emotional rapport, ultimately leading to increased satisfaction and achievement (Biocca et al., 2003). An AI avatar that responds to questions in real-time transforms the online lecture system from a mere information delivery tool into an interactive social entity, thereby potentially increasing the level of perceived social presence.

Conversely, according to CLT, extraneous cognitive load generated by unresolved inquiries or response delays results in the inefficient consumption of mental resources, which is detrimental to learning outcomes (Sweller, 1988; Mayer, 2021). The use of real-time Q&A is expected to reduce this unnecessary cognitive load, acting as a negative (–) mediation pathway (i.e., reducing a negative factor to improve performance).

Based on these theoretical underpinnings, this study proposes a mediation model in which the availability of real-time Q&A influences learning performance through the positive mediating effects of flow and social presence, and the negative mediating effect of cognitive load.

4.2 Main Effects of Response Speed and Linguistic Style

H3: Shorter response latency will lead to higher learner immersion and social presence, thereby enhancing learning performance.

According to Flow Theory, learners maintain a state of immersion when they receive immediate and unambiguous feedback during task execution. This immediacy is a key factor in increasing task concentration and the depth of information processing (Csikszentmihalyi, 1991). In an online lecture environment, the primary method for implementing such immediate feedback is by minimizing the system's response latency. Shorter response times allow learners to perceive the interaction as occurring in "real-time." This promptness encourages learners to view the avatar as a viable interaction partner and strengthens their positive evaluation of the online lecture system’s credibility and competence.

H4: A conversational style will enhance learning performance by increasing social presence and reducing cognitive load compared to a formal style.

Mayer’s (2021) Personalization Principle suggests that utilizing a conversational style in instructional delivery facilitates the learner’s attention and meaning-construction processes. By providing social cues that mimic human interaction, a conversational style activates the learner's social schemas, leading them to perceive the system not as a mere information conduit but as an interactive social entity (Mayer, 2021). This shift in perception is highly likely to increase the sense of social presence between the learner and the avatar.

Furthermore, Cognitive Load Theory posits that if the method of information presentation unnecessarily consumes a learner’s processing capacity, extraneous cognitive load increases, thereby hindering learning efficiency (Sweller, 1988). Formal expressions may cause learners to focus more on the linguistic format or the "documentlike" nature of the speech, potentially leading to cognitive distraction. Conversely, conversational expressions use familiar and accessible language, reducing the mental effort required to decode and understand the lecture content (Mayer, 2021). Consequently, it is hypothesized that the use of conversational language will provide greater social presence and lower cognitive burden, ultimately resulting in superior learning performance.

4.3 Interaction Effects between Response Speed and Linguistic Style

H5: There will be a significant interaction effect between response speed (fast vs. slow) and linguistic style (conversational vs. formal).

Mayer (2021) asserts that when instructional content is delivered in a conversational manner—as if engaging in a dialogue—learners perceive the system not merely as a tool for information retrieval but as a social partner for communication. From this perspective, the combination of a conversational style and fast response speed is expected to facilitate higher levels of learner engagement compared to traditional online environments. For instance, when a learner encounters a difficult concept and receives an immediate response such as, "That's a great question! Let me explain it again," it fosters a sense of "co-learning" rather than isolated study.

Such a synchronized environment is likely to reduce the cognitive burden associated with the learning process. Immediate resolution of inquiries eliminates the need for learners to pause the lecture or perform external internet searches, thereby preventing the disruption of flow. In contrast, significant latency in responses often leads learners to doubt the functional integrity of the system, prompting redundant actions such as refreshing the page or troubleshooting technical issues, which ultimately results in a loss of concentration.

Conversely, a system that responds rapidly preserves cognitive momentum and sustains focus naturally. When this technical efficiency is coupled with a conversational linguistic style, it elevates the AI avatar from a rigid explanatory tool to a social entity, significantly enhancing social presence. It is hypothesized that this synergy—where technical reliability (speed) meets social rapport (style)—will lead to a more profound perception of the avatar as a trustworthy and approachable partner, ultimately exerting a positive synergistic influence on learning performance.

5. Research Methodology

5.1 Research Design and Participant Recruitment

A total of 52 participants (N=52) were recruited for this study, comprising 18 males, 32 females, and 2 participants who declined to state their gender. The sample included undergraduate students, graduate students, and general adults ranging in age from their 20s to 40s. Regarding prior knowledge of the lecture topic—"Social Research Methodology"—19 participants (36.5%) reported previous learning experience, while 33 participants (63.5%) had no prior exposure. All experimental procedures were conducted in strict accordance with institutional research ethics and protocols. The study employed a 2 × 2 between-subjects factorial design, with the addition of a control group, resulting in a total of five experimental conditions. The two independent variables—response speed and linguistic style—were systematically manipulated within the four experimental subgroups to observe their individual and interactive effects on the dependent variables.

4-0-0 Qdrd`qbg Cdrhfm9Ehud,Fqnto Bnl o`q`shud Eq`l dvnqj

A total of five groups were established to evaluate four experimental conditions featuring the Q&A system against a baseline control group that received no Q&A functionality.

Table 1. Experimental group classification

Group Number	Group Type	Q&A Provision	Response Speed	Language Style
Control Group	Q&A Not Provided	Not Provided	N/A	N/A
Experimental Group 1	Q&A Provided (Fast/Conversational)	Provided	Fast (within 3.0 seconds on average)	Conversational
Experimental Group 2	Q&A Provided (Fast/Formal)	Provided	Fast (within 3.0 seconds on average)	Formal
Experimental Group 3	Q&A Provided (Slow/Conversational)	Provided	Slow (6.0 seconds or longer on average)	Conversational
Experimental Group 4	Q&A Provided (Slow/Formal)	Provided	Slow (6.0 seconds or longer on average)	Formal

4-0-1 O`qshbho`ms Qdbqthsl dms `mc @kknb`shnm

To ensure statistical rigor, a total of 52 participants were recruited and randomly assigned to one of the five experimental conditions. The collected data were analyzed to ascertain differences in learning outcomes and psychological constructs between the control group and the four experimental groups. Appropriate statistical methodologies were then employed to test the formulated hypotheses and evaluate the comparative effects of each interaction design.

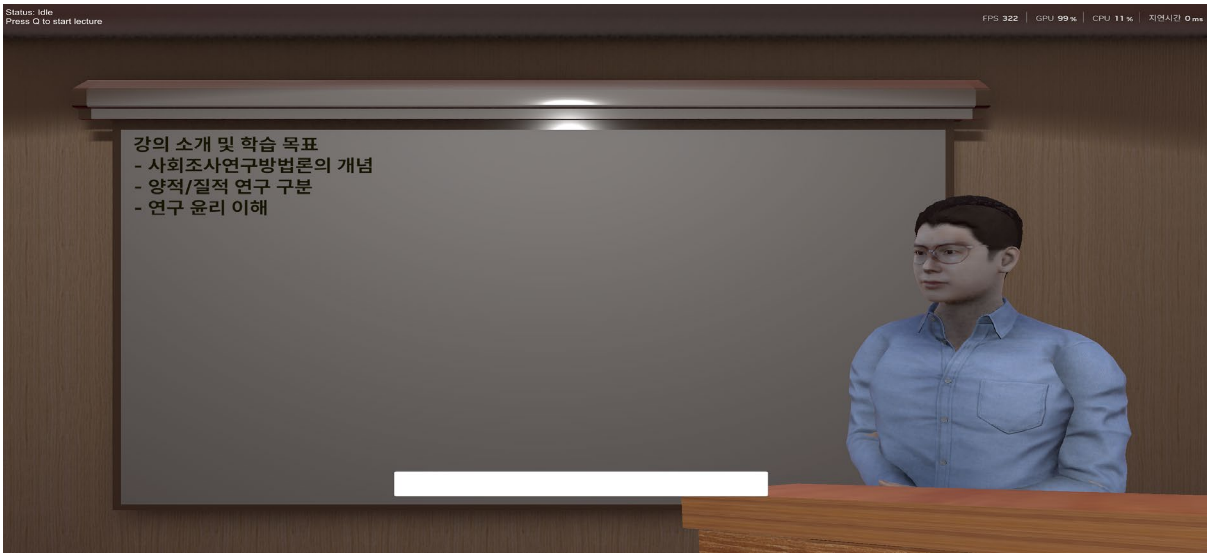
4-0-2 Dwodqhl dms`k Oqnbdct qd `mc L `mhot k`shnm

The online instructional system for this experiment was custom-developed using Unity 6000.2, integrated with a suite of artificial intelligence technologies. The lecture, titled "Social Research Methodology," was designed as a seven-minute video session. To facilitate multimodal interaction, the system allowed participants to engage in real-time Q&A by either utilizing voice-based inquiries via Speech-to-Text (STT) or text-based input through an interface located at the bottom of the screen (Figure 1). The technical architecture featured a high-fidelity digital avatar created through Reallusion Character Creator 5 (CC5), with vocal synthesis powered by ElevenLabs Text-to-Speech (TTS) and input processing handled by Google STT. For the inference engine, the Gemini-flash-latest Large Language Model (LLM) was deployed; the "Flash" model was specifically selected to ensure the rapid inference speeds necessary for seamless real-time interaction without compromising reasoning accuracy.

The experimental design systematically manipulated two independent variables: response speed and linguistic style. Response speed was operationalized by maintaining an average latency of under 3.0 seconds for the "Fast" condition, while a delay of over 6.0 seconds was artificially introduced for the "Slow" condition. Linguistic style was clearly bifurcated into a conversational style (utilizing the Korean haeyo-che) to encourage rapport and a formal style (utilizing hapshyo-che) to maintain pedagogical decorum.

Prior to the main study, a pilot test with ten participants led to significant refinements in the experimental environment. To mitigate potential voice recognition errors, the chat interface was enhanced, and concise on-screen summaries were added to support content comprehension (Figure 1). The validity of these manipulations was objectively verified through system logs, which confirmed that the actual response times adhered strictly to the intended experimental design (under 3s for the fast condition and over 6s for the slow condition). Linguistic consistency was ensured through rigorous prompt engineering and expert pre-review. While participants were informed of these system characteristics, it should be noted that a separate subjective survey for manipulation checks based on participant perception was not administered.

Figure 1. Unity Experimental Environment



5.2 Manipulation and Implementation of Independent Variables

4-1-0 Dwodqhldms`k Bnmchshnmr A`rdc nmP %@@u`hk`ahkhsx

The participants were categorized into either a control group or one of four experimental treatment groups based on the availability of interactive functionality during the learning process. Participants in the control group were placed in a passive viewing environment that mirrors traditional asynchronous online learning, where no real-time Q&A functionality was provided. To simulate a realistic learning scenario, these participants were explicitly instructed prior to the session to pause the video and perform external information seeking—such as internet searches or utilizing smart devices—should they encounter any ambiguities or questions regarding the lecture content.

In contrast, participants in the experimental groups engaged in a learning environment facilitated by a real-time Q&A system. In this condition, learners could pose spontaneous inquiries (e.g., "Could you explain qualitative research in more detail?"), which were then recognized and addressed vocally by the AI avatar. As established in the research design, these experimental groups were further subdivided into four conditions based on the 2×2 factorial arrangement of response speed (fast vs. slow) and linguistic style (conversational vs. formal). To minimize potential confounding variables related to technical unfamiliarity, participants in the experimental groups underwent a pre-experimental orientation session, where they practiced voice-mediated inquiries to ensure they were fully proficient with the interaction process before the actual data collection commenced.

4-1-1 L`mhotk`shnm ne Qdronmrd K`sdmbx

In this study, response latency is defined as the temporal interval from the moment the system recognizes a learner’s inquiry to the initiation of the AI avatar's vocal output. This variable was operationalized into two distinct conditions: Fast and Slow.

The Fast condition was designed to provide an average response within 3 seconds. To achieve this near-instantaneous interaction, the system utilized the Google Gemini-flash-latest LLM, known for its rapid inference capabilities. Furthermore, to minimize perceived waiting time, a Python-based processing technique was employed to segment generated text into smaller sentence units. These units were then immediately converted via TTS and played sequentially, simulating the flow of a real-time conversation. This duration mirrors typical synchronous messaging environments where a partner responds immediately, allowing the learner to maintain cognitive momentum and perceive the interaction as a seamless dialogue rather than a disjointed technical process.

Conversely, the Slow condition was manipulated to ensure an average delay of 6 seconds or more before the avatar initiated a response. While 6 seconds may appear brief in isolation, it represents a significant duration in the context of human-computer interaction, often exceeding the threshold for maintaining a "sense of presence." During this interval of inactivity, learners may experience cognitive friction, leading to doubts regarding the system’s functional integrity or their own input accuracy. Such delays are hypothesized to disrupt the "concentration on the task" component of flow, potentially diverting the learner’s attention toward the technology itself rather than the instructional content.

By contrasting these two conditions, this study aims to empirically evaluate how the temporal dimension of AI-mediated interaction reconfigures learner immersion, perception of the avatar, and the overall educational experience.

4-1-2 L `mhot k`shnmne Khmf t hrshb Rsxkd

In this study, linguistic style refers to the register, tone, and discursive patterns employed by the AI avatar during both the instructional delivery and the Q&A sessions. Grounded in Mayer's (2021) Personalization Principle, the avatar's linguistic persona was bifurcated into two distinct conditions: Conversational and Formal.

The Conversational condition was designed to minimize psychological distance and foster social immediacy between the learner and the avatar. This condition utilized the Korean haeyo-che (a polite yet friendly and informal honorific style), characterized by sentence endings such as "-yo" or "-eyo." For instance, when explaining triangulation, the avatar would respond: "It's a method to increase the reliability of research results by analyzing a single object or phenomenon through various methods and data simultaneously." The objective of this style was to create a comfortable, interactive atmosphere that encourages the learner to perceive the avatar as a helpful social partner.

Conversely, the Formal condition focused on the authoritative and official delivery of information. This condition employed the hapshyo-che (a highly formal and professional honorific style), utilizing sentence endings such as "-nida" or "-imnida." For the same inquiry, the avatar would provide an identical semantic explanation but with a stiff, business-like tone: "This is a method intended to enhance the reliability of research findings by concurrently analyzing a single subject via multiple methodologies." This style reflects the conventional tone found in textbooks or formal academic lectures, emphasizing professional decorum over personal rapport.

By isolating these linguistic markers while maintaining identical core content, this study seeks to observe how subtle variations in speech style reconfigure the learner's perception of familiarity and the overall quality of the interaction.

5.3 Measurement of Dependent and Mediating Variables

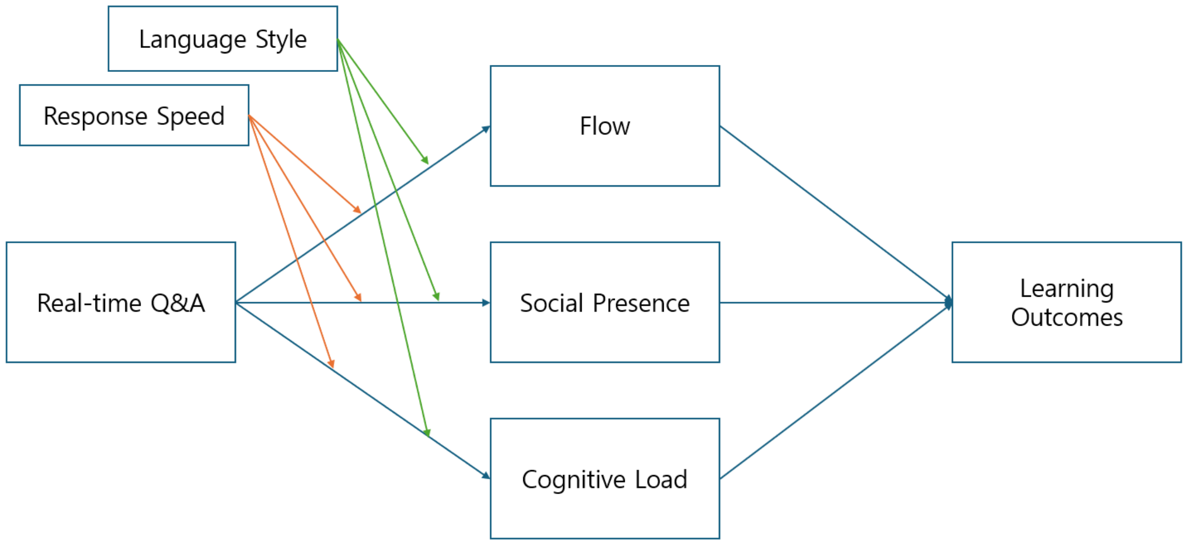
Immediately following the experimental session, participants completed a comprehensive survey to assess their learning experience. The primary psychological constructs—flow, social presence, and cognitive load—were measured using validated self-report instruments. Each item was rated on a 5-point Likert scale, ranging from "1: Strongly disagree" to "5: Strongly agree." This allowed for a quantitative assessment of internal states, such as the degree of concentration or the perceived social connection with the avatar.

Objective learning performance was evaluated through a pre-test and post-test knowledge assessment rather than subjective ratings. Both tests were designed with equivalent difficulty levels to measure the gain in knowledge. By calculating the difference between pre- and post-test scores, the study aimed to verify the actual pedagogical efficacy of the system beyond mere learner satisfaction.

Furthermore, to complement the quantitative data, a single open-ended question was included. This allowed participants to freely describe their subjective experiences and provide qualitative insights into the AI-based lecture environment that might not be fully captured by numerical scales.

Figure 2 illustrates the integrated research model. The model depicts the structural relationships where the availability of real-time Q&A and its associated variables (response speed and linguistic style) influence the mediating constructs—flow, social presence, and cognitive load—which in turn affect terminal learning performance. By triangulating Likert-scale data with qualitative feedback, this study sought to provide a granular analysis of how specific interaction designs reconfigure the learner experience.

Figure 2. Research model



5.4 Data Analysis Methods

To comprehensively evaluate the experimental results, this study analyzed data from five distinct groups: one control group (no Q&A functionality) and four experimental groups structured within a 2 × 2 factorial design (Response Speed × Linguistic Style).

Data curation and preprocessing were initially performed using Microsoft Excel. All subsequent statistical procedures, including descriptive and inferential analyses, were conducted using IBM SPSS Statistics (Version 31.0.1.0). The analysis focused on identifying significant differences and interaction effects across the groups to validate the proposed research model and hypotheses.

5.4.1 Verification of the Effect of Q&A Provision (H1)

To test Hypothesis 1 (H1), which posits a significant difference in learning performance based on the availability of a real-time Q&A system, the participant pool was reclassified. Participants were categorized into two primary groups: the control group, which received no Q&A support, and the treatment group, which integrated the four experimental subgroups that utilized the system.

An independent samples t-test was conducted to statistically examine the mean difference in learning performance between these two categories. Additionally, a one-way analysis of variance (ANOVA) was performed to provide a more comprehensive assessment of inter-group variability and to validate the baseline comparison. This dual-analytical approach was intended to multidimensionally verify the existence of a main effect exerted by the real-time Q&A system on learning outcomes.

4-3-1 @m`kxrhr ne L`hm `mc Hmsdq`bshnm Deedbsr 'G2+G3+`mc G4(

To examine the specific impact of system characteristics on learning performance and psychological constructs in greater detail, the control group was excluded from this stage of the analysis, focusing exclusively on the four experimental subgroups. These subgroups were differentiated by the 2×2 factorial manipulation of response speed (Fast vs. Slow) and linguistic style (Conversational vs. Formal).

A 2×2 two-way analysis of variance (ANOVA) was conducted to verify the main effects of response speed (H3) and linguistic style (H4), as well as the potential interaction effect between the two variables (H5). This analytical approach was utilized to identify not only the independent influence of each factor but also how their combination synergistically reconfigures the learner's cognitive and social experience.

4-3-2 Udqhehb`shnm ne Ldch`shnf Deedbsr 'G1(

To test Hypothesis 2 (H2), which examines the indirect effects of the independent variables (Q&A availability, response speed, and linguistic style) on learning performance via psychological mediators (flow, social presence, and cognitive load), a bootstrapping procedure was employed.

Bootstrapping was selected for its distinct advantage in not requiring the assumption of a normal distribution, making it a robust analytical tool for obtaining stable estimates even with relatively small sample sizes, as in the present study. Furthermore, it is widely recognized as a highly appropriate method for verifying the significance of indirect effects that operate through complex mediating pathways.

The results were interpreted based on 95% bias-corrected confidence intervals (CIs). The significance of the mediating effects was determined by verifying whether zero (0) was included within the calculated confidence interval. Following standard inferential protocols, if the interval between the lower and upper bounds does not contain zero, the mediation effect is deemed to be statistically significant.

6. Experimental Results

This chapter reports the quantitative and qualitative findings regarding the impact of interaction modalities in AI avatar-based online lectures on learning performance and user experience. The quantitative data analyzed herein constitute learning performance (ScoreDiff), measured by the gain between pre- and post-test knowledge assessments, and various psychological and experiential constructs derived from self-report surveys, specifically flow (FlowAvr), social presence (SocialPresenceAvr), cognitive load (CognitiveLoadAvr), satisfaction (SatisfactionAvr), subjective achievement (GradeFeelAvr), and overall user experience (UXAvr). To complement these numerical findings, qualitative data were obtained by segmenting open-ended responses into semantic units and subsequently categorizing them into thematic results.

The chapter is structured to provide a comprehensive analysis of the research hypotheses. Section 6.1 begins by evaluating the effect of real-time Q&A availability on learning performance (H1). Subsequently, Section 6.2 focuses on the Q&A-enabled conditions to examine the main and interaction effects of response speed and linguistic style on both learning outcomes and affective experiences (H3–H5) using a two-way ANOVA. Section 6.3 then details the results of the mediation and moderated mediation analyses (H2), followed by Section 6.4, which provides a granular report of descriptive statistics and survey patterns across all experimental conditions. In Section 6.5, the qualitative feedback is expanded to triangulate and bolster the observed quantitative patterns. Finally, Section 6.6 synthesizes both the quantitative and qualitative data to summarize the core implications of the experimental results.

6.1 Verification of the Effect of Real-time Q&A Provision (H1)

5-0-0 C`s` Rsqtbstqd `mc U`qh`akd Cdehmhshnm

The study sample comprised a total of $N = 52$ participants, distributed across one control group and four experimental treatment groups. The control group ($n = 11$) served as a baseline condition representing a conventional online lecture environment where participants viewed video content without real-time Q&A functionality. In contrast, the experimental groups were structured according to a 2×2 factorial arrangement that manipulated response speed (Fast vs. Slow) and linguistic style (Conversational vs. Formal). These conditions were categorized into four distinct subgroups: Group B (Conversational-Fast), Group C (Formal-Fast), Group D (Conversational-Slow), and Group E (Formal-Slow). The sample sizes were relatively balanced across the conditions, with each group consisting of 10 to 11 participants.

To evaluate the efficacy of the intervention, learning performance was operationalized as ScoreDiff, a key dependent variable representing the improvement in academic achievement. This was calculated by subtracting the pre-test scores ($PreTest$) from the post-test scores ($PostTest$), where $ScoreDiff = PostTest - PreTest$. Additionally, the affective and experiential dimensions of the learning process were summarized using composite indices derived from self-report surveys. These indices encompassed various psychological constructs, specifically flow, social presence, cognitive load, lecture satisfaction, subjective achievement ($GradeFeel$), and overall UI/UX satisfaction.

5-0-1 Cdrbqhoshud Rs`shrshbr

An examination of the group means revealed that the control group's score increased by 0.55, rising from a pre-test mean of 12.45 to a post-test mean of 13.00. In contrast, all experimental conditions exhibited mean gains ranging from 0.64 to 1.50. While the Q&A-enabled groups demonstrated a directional trend toward greater improvement compared to the control group, the statistical significance of these differences requires verification through inferential statistics. The comprehensive descriptive statistics for all five groups are consolidated in Table 2 and Table 3.

The descriptive data highlight two significant patterns of divergence. First, regarding learning performance (ScoreDiff), Group E demonstrated the most substantial improvement ($M = 1.50$), followed by Group B ($M = 1.10$). Second, concerning the affective experience — encompassing flow, social presence, cognitive load, satisfaction, and subjective achievement—Group B displayed the most favorable overall profile. Specifically, Group B reported the highest levels of flow ($M = 4.08$) and social presence ($M = 3.68$), while maintaining the lowest cognitive load ($M = 1.97$), suggesting a superior quality of learner experience in this condition. Furthermore, Group B achieved the highest scores in both lecture satisfaction ($M = 4.20$) and subjective learning achievement ($M = 4.26$).

Notably, the results indicate that the condition yielding the greatest objective achievement gain (Group E) did not necessarily coincide with the condition providing the most optimal psychological experience (Group B). This divergence suggests that significant improvements in learning outcomes do not always equate to the highest levels of perceived learner engagement or satisfaction. These discrepancies offer critical implications for the design of interactive online learning environments and will be discussed in-depth in Chapter 7.

5-0-2 Hmcdodmcdms R`l okdr t,sdrs

Hypothesis 1 (H1) posits that the availability of a real-time Q&A system significantly influences learning performance. To test this hypothesis, the control group (Group A) was compared against the pooled experimental groups (Groups B–E) to examine the mean differences in ScoreDiff.

The group statistics are summarized as follows: the control group ($N = 11$) yielded a mean improvement of $M = 0.5455$ with a standard deviation of $SD = 5.1257$, while the integrated experimental group ($N = 41$) showed a mean gain of $M = 0.9756$ with a standard deviation of $SD = 2.1271$. Given the substantial discrepancy in the standard deviations between the two groups, the assumption of homogeneity of variance (homoscedasticity) was rigorously examined using Levene's test. This preliminary check is essential to determine whether to report the t-statistic assuming equal variances or the adjusted version (Welch's t-test) to ensure the statistical validity of the comparison.

Table 2. Descriptive Statistics1 for Each Group

Experimental Groups	N	PreTest	PostTest	ScoreDiff
Control Group (A: No Q&A)	11	12.45	13.00	0.55
Group B (Conversational Tone – Fast Response)	10	13.30	14.40	1.10
Group C (Formal Tone – Fast Response)	10	13.20	13.90	0.70
Group D (Conversational Tone – Slow Response)	11	14.55	15.18	0.64
Group E (Formal Tone – Slow Response)	10	14.10	15.60	1.50

Levene's test for equality of variances indicated that the assumption of homogeneity of variance was violated ($p = .001$). Consequently, the t-test results not assuming equal variances (Welch's t-test) were utilized to ensure statistical accuracy. The analysis revealed no statistically significant difference in ScoreDiff between the control group ($M = 0.55$) and the integrated experimental group ($M = 0.98$), with $t(10.940) = -0.272$, $p = .791$ (two-tailed). These results suggest that the mere availability of a real-time Q&A system did not lead to a statistically superior gain in objective learning performance compared to the control condition. Therefore, Hypothesis 1 (H1) was not supported.

Table 3. Descriptive Statistics2 for Each Group

Experimental Groups	Flow	Social Presence	Cognitive Load	Satisfaction	Grade Feel	UX
Control Group (A: No Q&A)	3.29	3.04	2.73	3.82	3.18	3.78
Group B (Conversational Tone – Fast Response)	4.08	3.68	1.97	4.20	4.26	3.88
Group C (Formal Tone – Fast Response)	3.86	2.98	2.53	4.00	3.72	3.38
Group D (Conversational Tone – Slow Response)	3.91	2.96	2.58	4.00	3.67	3.60
Group E (Formal Tone – Slow Response)	3.68	3.44	2.40	3.90	3.78	3.54

These findings suggest that, within the current sample, there is insufficient empirical evidence to conclude that the mere provision of a real-time Q&A system significantly enhances ScoreDiff. However, a descriptive analysis reveals a noteworthy trend: the integrated experimental group exhibited a higher mean improvement compared to the control group ($0.9756 > 0.5455$).

More importantly, a marked discrepancy was observed in the dispersion of scores between the two conditions. While the control group demonstrated high volatility in learning outcomes ($SD = 5.1257$), the experimental groups showed a much more condensed distribution ($SD = 2.1271$). This indicates that while the system did not yield a statistically significant "mean-shifting" effect, it may have played a critical role in reducing the variability of learning achievement. In other words, the real-time Q&A functionality might function as a stabilizing intervention that fosters more consistent learning gains across diverse learners. This observation regarding "pedagogical stability" provides a significant basis for further discussion in Chapter 7 concerning the design of reliable educational environments (Table 4, Table 5).

Table 4. Group Statistics and Independent Samples t-Test for the Effect of Q&A Availability

	Statistics				
	QA_Group	**N**	**MEAN**	**SD**	**SE**
Score Diff	Control group	11	.5455	5.12569	1.54545
	Experimental group	41	.9756	2.12706	.33219

Table 5. Independent Samples t-Test for SDiff

Test	Value

Levene	Levene’s F	11.562
	Levene’s Sig	.001
t-test	t	-.272
	Df	10.940
	One-tailed p	.395
	Two-tailed p	.791
	Mean Difference	-.430
	Std.Error Differernce	1.581
95% **CI**	95% CI Lower	-3.912
	95% CI Uppder	3.051

* *Note.* Equal variances not assumed based on Levene's test (F = 11.562, p = .001)

5-0-3 Cdbnlonrhshmne RbnqdChee ax Oqhnq Kd`qmhmf Dwodqhdmbd

Table 6 presents the decomposition of the mean ScoreDiff based on the presence or absence of prior learning experience in the "Social Research Methodology" course. The results reveal distinct mean patterns across conditions depending on the learners' background knowledge. In this analysis, $PreStudyExp = 0$ denotes the group with no prior experience, while $PreStudyExp = 1$ refers to the group with prior experience.

For learners without prior experience ($PreStudyExp = 0$), the mean ScoreDiff for each group was as follows: the Control group = 0.50 ($n = 8$), Condition B (Conversational–Fast) = 1.17 ($n = 6$), Condition C (Formal–Fast) = 0.14 ($n = 7$), Condition D (Conversational–Slow) = 0.14 ($n = 7$), and Condition E (Formal–Slow) = 1.00 ($n = 5$). Within this segment, Condition B exhibited the largest mean increase, followed by Condition E, whereas the gains in Conditions C and D were relatively limited.

Conversely, for learners with prior experience ($PreStudyExp = 1$), the mean ScoreDiff across groups was: the Control group = 0.67 ($n = 3$), Condition B = 1.00 ($n = 4$), Condition C = 2.00 ($n = 3$), Condition D = 1.50 ($n = 4$), and Condition E = 2.00 ($n = 5$). In this decomposition, Conditions C (Formal–Fast) and E (Formal–Slow) showed the highest mean values (2.00 each), followed by Condition D (1.50).

These patterns suggest that the efficacy of a specific "interaction persona"—the combination of response speed and linguistic style—may be contingent upon the learner's prior knowledge level. Specifically, while the Conversational–Fast condition (B) appeared more effective for novice learners, the Formal conditions (C and E) tended to be associated with greater achievement gains for experienced learners.

However, as the sample size for each cell is exceedingly small, ranging from a minimum of 3 to a maximum of 7 participants, there are inherent limitations in presenting these results as statistically definitive conclusions. Consequently, this section characterizes these findings solely as an exploratory "observed tendency." The necessity for adaptive interaction designs tailored to learner levels is discussed further in Section 7.5.

Table . Descriptive Statistics for ScoreDi楊by Experimental Group and Prior Learning Experience

ExpGroup	PreStudyExp	MEAN	SD	N
Control Group	No Prior Learning Experience	.	.	
	With Prior Learning Experience	. 歯	.歯	
	Total	.	.	
Conversational Tone – Fast Response	No Prior Learning Experience	. 歯	.	
	With Prior Learning Experience	.	.	
	Total	.	.歯 歯	
Formal Tone – Fast Response	No Prior Learning Experience	.	.	歯
	With Prior Learning Experience	.	.歯	
	Total	.歯	. 歯	
Conversational Tone – Slow Response	No Prior Learning Experience	.	.	歯
	With Prior Learning	.	. 歯	

	Experience			
	Total	.	.	
Formal Tone – Slow Response	No Prior Learning Experience	.	.	
	With Prior Learning Experience	.	.歯歯	
	Total	.	. 歯	
Total	No Prior Learning Experience	.	.	
	With Prior Learning Experience	. 歯歯	. 歯	
	Total	.	.	

5-0-4 Nt skhdq Rdmrhshuhsx `mc Qnatrsmdrr Bgdbj

To ensure the empirical validity of the findings and minimize the risk of mean distortion or violations of homoscedasticity, a robustness check was conducted. Specifically, extreme values in ScoreDiff ($-8, 6, 7, 11$) were identified as outliers and excluded from the dataset, after which the mean pre- and post-test scores for each group were recalculated (Table 7).

After removing these outliers, the pre-test means were 13.75 for Group A (Control), 13.89 for Group B, 13.20 for Group C, 14.55 for Group D, and 14.10 for

Group E. The corresponding post-test means were recorded as 13.38 for Group A, 14.33 for Group B, 13.90 for Group C, 15.18 for Group D, and 15.60 for Group E. In terms of mean gain, the control group showed a slight decrease of -0.37, whereas all experimental groups (B–E) exhibited gains ranging from 0.44 to 1.50. These results indicate that even when extreme cases are excluded, the groups utilizing the real-time Q&A system consistently demonstrated more positive trends in learning achievement. Notably, Group E maintained the largest mean improvement (1.50), and Group B likewise sustained a positive trajectory without any reduction in scores.

However, it is important to note that the removal of outliers further reduced the sample size to between 8 and 11 participants per cell (Control: $n = 8$; B: $n = 9$; C: $n = 10$; D: $n = 11$; E: $n = 10$). Consequently, these findings possess inherent limitations for being interpreted as definitive statistical effects. Instead, this analysis should be viewed as a robustness check that confirms the directional stability of the results, providing supplementary evidence that supports the primary main effect analysis.

Table 歯 Group-wise Pre-test and Post-test Means and Di楊erences

Group	Pre-test Mean	Post-test Mean	Di楊erence (Post – pre)
Control Group (A: No Q&A)	.歯	.	- . 歯
Group B (Conversational Tone – Fast Response)	.	.	.
Group C (Formal Tone – Fast Response)	.	.	.歯
Group D (Conversational Tone	.	.	.

– Slow Response)			
Group E (Formal Tone – Slow Response)	.	.	.

5-1 Udqhehb`shnm ne L `hm `mc Hmsdq`bshnm Deedbsr ne Qdronmrd Roddc `mc Khmf thrshb Rsxkd 'G2"G4(

Focusing exclusively on the conditions where real-time Q&A functionality was provided ($N = 41$), this section evaluates the main effects of Response Speed (Fast vs. Slow) and Linguistic Style (Conversational vs. Formal), as well as their potential interaction effects. The statistical analysis was conducted using a series of 2×2 two-way analyses of variance (ANOVA). For each dependent variable, the findings are reported in the following systematic sequence: (1) descriptive statistics, including means and standard deviations; (2) Levene's test for the equality of error variances; (3) inferential statistics, including F-statistics, p-values, and partial eta squared (η_p^2); and (4) a comprehensive interpretation focusing on statistical significance and effect size.

5-1-0 Sv n,v `x @MNU@ enq Kd`qmhmf Odqenq l `mbd 'RbnqdChee(

Prior to conducting the 2×2 two-way ANOVA, a one-way ANOVA was performed to assess the homogeneity of pre-test scores across groups. Since the assumption of homogeneity of variance was violated—Levene's test: $F(4,47) = 7.54, p < .001$—Welch's ANOVA was applied as a robust alternative. The results indicated no statistically significant differences in $PreTest$ scores among the five groups—$F(4,22.5) = 0.906, p = .477$—with the control group reporting the lowest mean (12.5) and the conversational-slow group reporting the highest (14.5). This confirms that the baseline knowledge levels were equivalent across all conditions prior to the experimental intervention.

To examine the main and interaction effects of response speed and linguistic style on the improvement of learning achievement, a 2×2 two-way ANOVA was performed on $ScoreDiff$ (Table 9). Preliminary testing confirmed that the assumption of homogeneity of variance was met—Levene's test: $F(3,37) = 1.09, p = .362$ (Table 10). The descriptive statistics for each condition were as follows: Group B (Conversational-Fast, $n = 10, M = 1.10, SD = 2.77$), Group C (Formal-Fast, $n = 10, M = 0.70, SD = 2.45$), Group D (Conversational-Slow, $n = 11, M = 0.64, SD = 2.01$), and Group E (Formal-Slow, $n = 10, M = 1.50, SD = 1.18$) (Table 8).

The two-way ANOVA revealed that the main effect of response speed was not statistically significant—$F(1,37) = 0.061, p = .807, \eta_p^2 = .002$. Similarly, the main effect of linguistic style was not significant—$F(1,37) = 0.116, p = .736, \eta_p^2 = .003$. Furthermore, the interaction effect between response speed and linguistic style did not reach statistical significance—$F(1,37) = 0.859, p = .360, \eta_p^2 = .023$ (Table 9). In terms of effect size, the main effects were negligible, and the interaction effect was small by standard criteria.

Despite the lack of statistical significance, the descriptive patterns exhibited a distinct disordinal (crossing) interaction. Under the conversational style, fast responses (B) led to higher gains than slow responses (D); conversely, under the formal style, slow responses (E) outperformed fast responses (C). This crossing pattern suggests that learning gains may not be driven by individual factors alone, but rather by the contextual congruence between the interaction rhythm and the linguistic persona. This observation, alongside the marginal interaction effect observed in social presence ($p = .051$), will be expanded upon in Chapter 7 to discuss the importance of cue-alignment (linguistic vs. temporal cues) in AI avatar design.

Table 8. Descriptive Statistics (Means, Standard Deviations) and Cell Sample Sizes for ScoreDiff in the 2 × 2 Design (Dependent Variable: ScoreDiff)

Language Style	**Response Speed: Fast** **Mean(SD), N**	**Response Speed: Slow** Mean(SD), N
Conversational Tone	1.10 (2.77), n=10	0.64 (2.01), n=11
Formal Tone	0.70 (2.45), n=10	1.50 (1.18), n=10

Table 9. Summary of the 2 × 2 Two-Way ANOVA Results for ScoreDiff

Effect	**F(1,37)**	**p**	**Partial Eta Squared (ηp^2)**
Response Speed (Fast vs. Slow)	0.061	0.807	0.002
Language Style (Conversational vs. Formal)	0.116	0.736	0.003
Response Speed × Language Style	0.859	0.360	0.023

Table 10. Results of Levene's Test for Homogeneity of Variance

Test	**p**
Levene's Test for Homogeneity of Variances	0.362

Figure 3. Estimated Marginal Means of Score Difference(PostTest – PreTest)

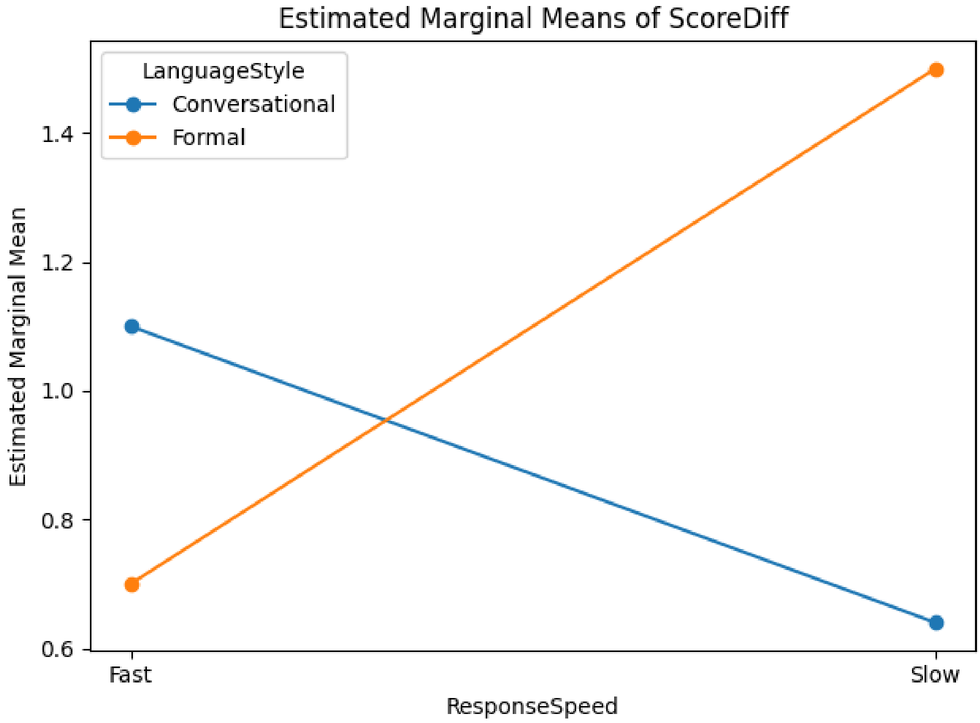


Figure 3 illustrates the interaction patterns of response speed and linguistic style regarding learning performance (ScoreDiff). The two-way ANOVA results indicated that neither the main effect of response speed ($F(1,37) = 0.061, p = .807$) nor the main effect of linguistic style ($F(1,37) = 0.116, p = .736$) was statistically significant. Furthermore, the interaction effect between the two variables did not reach the significance threshold ($F(1,37) = 0.859, p = .360$) (Table 9).

However, an examination of the estimated marginal means reveals distinct, contrasting patterns across conditions. In the conversational condition, fast responses yielded higher mean achievement gains compared to slow responses (Group B: 1.10 vs. Group D: 0.64). Conversely, in the formal condition, slow responses resulted in higher achievement gains than fast responses (Group E: 1.50 vs. Group C: 0.70). This suggests that the impact on learning outcomes is not dictated by a single isolated factor but is rather contingent upon how these two interactional cues are synthesized.While the interaction effect was not statistically significant, its effect size ($\eta_p^2 = .023$) was notably larger than the effect sizes of the individual main effects (η_p^2 ranging from $.002$ to $.003$). This discrepancy implies that with a larger sample size, the interaction between linguistic and temporal cues might have manifested more definitively.

From a design perspective, these results suggest the absence of a universal "optimal" response modality for AI avatar tutors. For instance, a conversational-fast approach may be more suitable for scenarios where learner flow and immediate engagement are paramount. In contrast, a formal-slow approach may be more effective for delivering professional content or explaining complex concepts that require deliberate contemplation.

Through the lens of the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986), the formal style may have functioned as a social frame that prompted learners to adopt a more serious and analytical posture. According to ELM, when learners process information through the central route, they engage in a higher level of cognitive elaboration, leading to better internalization of knowledge. While the conversational style fosters a familiar and comfortable rapport, the formal style may activate a psychological mechanism that encourages learners to treat the lecture content with more gravity, thereby facilitating deeper analytical processing.

Furthermore, applying Cognitive Load Theory (Sweller, 1988), the latency inherent in the "slow" response condition may not have been a mere void, but rather a period for spontaneous reflection and cognitive restructuring. While immediate feedback in the conversational-fast condition is advantageous for sustaining flow (Csikszentmihalyi, 1991), the waiting period in the formal-slow condition may have allowed learners to mentally rehearse the lecture material and develop associative thoughts. The synergy between the serious reception attitude induced by the formal style and the reflective expansion triggered by the slower pace likely contributed to the observed achievement gains. However, as this study did not directly measure reflective behavior or depth of processing, further empirical validation is required.

These interpretative threads align with the marginal interaction trend observed in social presence ($p\ =\ .051$), reinforcing the premise that the congruence between linguistic expression and response speed significantly shapes learning performance. A more detailed discussion of these implications will be provided in Chapter 7.

5-1-1 Sv n, v `x @MNU@ enq Eknv 'Eknv @uq(

A $2\ \times 2$ two-way analysis of variance (ANOVA) was conducted to examine the main and interaction effects of response speed (fast vs. slow) and linguistic style (conversational vs. formal) on flow (FlowAvr). Prior to the analysis, the assumption of homogeneity of variance was verified using Levene's test, which yielded a non-significant result ($p\ =\ .421$), confirming that the assumption was satisfied.

The sample comprised four experimental conditions with the following sizes: conversational-fast ($n = 10$), formal-fast ($n = 10$), conversational-slow ($n = 11$), and formal-slow ($n = 10$). The mean flow scores for each group were recorded as 4.08 (Group B), 3.86 (Group C), 3.91 (Group D), and 3.68 (Group E), with standard deviations of 0.83, 0.48, 0.70, and 0.73, respectively.

The two-way ANOVA results indicated that the main effect of response speed was not statistically significant—$F(1{,}37) = 0.645, p = .427, \eta_p^2 = .017$—nor was the main effect of linguistic style—$F(1,37) = 1.057, p = .311, \eta_p^2 = .028$. Furthermore, the interaction effect between response speed and linguistic style was virtually non-existent and statistically non-significant ($F(1,37) = 0.000, p = .984, \eta_p^2 = .000$). Given that the effect sizes (η_p^2 ranging from $.017$ to $.028$) were small and the interaction effect was negligible, these findings suggest that the combination of linguistic style and response speed did not produce a synergistic impact on flow within the parameters of this sample.

As shown in Table 11, the conversational style conditions exhibited a trend toward higher mean flow scores compared to the formal style conditions. Specifically, conversational groups reported higher immersion than formal groups under both fast and slow response conditions. This pattern implies that learners may experience greater immersion when interacting via a conversational style, regardless of response latency. However, as these differences did not reach statistical significance, the results may have been influenced by the limited sample size or individual participant variance. While individual fluctuations were assumed to be balanced across groups when calculating means, this study draws the limited conclusion that although a "style effect" may be present, the current data are insufficient to confirm it definitively.

Beyond interactional immediacy, flow can be influenced by a multitude of confounding variables, including content difficulty, prior learner interest, the viewing environment, speech synthesis quality, and the avatar's non-verbal cues such as facial expressions and gestures. Qualitative feedback, which highlighted unnatural prosody and stiff avatar movements as barriers to immersion, suggests that the latent effects of linguistic style and response speed may have been attenuated by these noise factors.

This interpretation underscores the necessity for rigorous multivariate control of flow determinants in future research, as further discussed in Section 7.5

Table 11. 2 × 2 Cell Descriptive Statistics for Flow

Language Style	Response Speed: Fast Mean(SD), N	Response Speed: Slow Mean(SD), N
Conversational Tone	4.08 (0.83), n = 10	3.91 (0.70), n = 11
Formal Tone	3.86 (0.48), n = 10	3.68 (0.73), n = 10

Table 12. Summary of the 2 × 2 Two-Way ANOVA Results for Flow

Effect	F(1,37)	p	Partial Eta Squared (ηp^2)
Response Speed (Fast vs. Slow)	0.645	0.427	0.017
Language Style (Conversational vs. Formal)	1.057	0.311	0.028
Response Speed × Language Style	0.000	0.984	0.000

Table 13. Levene's Test for Flow

Test	p
Levene's Test for Homogeneity of Variances	0.421

Figure 4. Estimated Marginal Means of Flow

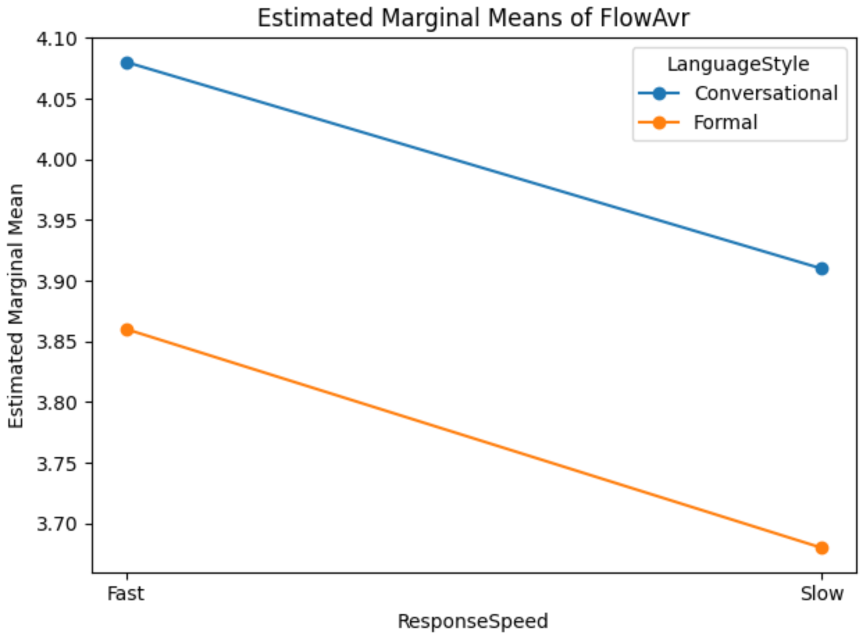


Figure 4 visualizes the estimated marginal means of learner flow according to response speed and linguistic style. As illustrated in the graph, both linguistic style conditions exhibited a similar downward trend, where flow levels were higher at faster response speeds and decreased slightly as response latency increased. In the conversational condition, the mean flow score was highest at the fast response speed ($M = 4.08$) and declined to 3.91 for slow responses. Similarly, the formal condition yielded scores of 3.86 and 3.68 for fast and slow responses, respectively, maintaining the same directional shift observed in the conversational condition.

Notably, the near-parallel orientation of the lines for both conditions visually indicates that the effect of response speed did not significantly vary by linguistic style. This suggests that the combination of these two variables likely lacks a synergistic interaction effect that either amplifies or inhibits flow, which is consistent with the two-way ANOVA results showing an interaction effect near zero ($F(1,37) = 0.000, p = .984$).

However, focusing on mean levels alone, it is noteworthy that the conversational style condition maintained consistently higher flow scores than the formal style across all response speed intervals. The conversational style outperformed the formal style in both the fast response condition ($4.08\ vs. 3.86$) and the slow response condition ($3.91\ vs. 3.68$). While not statistically significant, these patterns suggest that a friendly linguistic style may have an overall upward-adjusting effect on learners' subjective flow experiences.

In summary, Figure 4 visually demonstrates that while the combination of response speed and linguistic style did not form a structural interaction effect on flow, conversational-based interaction showed a weak directional trend associated with higher immersion levels. This supports the premise that flow is a construct likely influenced by a complex array of factors beyond simple interactional cues—such as content difficulty, vocal/avatar quality, and learning context—and implies that these confounding noise factors in the experimental environment may have attenuated the latent effects of linguistic style and response speed.

5-1-2 Sv n,v `x @MNU@ enq Rnbh`k Oqdrdmbd 'Rnb@uq(

A 2×2 two-way analysis of variance (ANOVA) was conducted to evaluate the main and interaction effects of response speed (fast vs. slow) and linguistic style (conversational vs. formal) on social presence (SocAvr). Prior to the analysis, Levene's test for equality of error variances yielded a non-significant result, confirming that the assumption of homogeneity of variance was satisfied. The sample sizes and mean scores for each condition were as follows: conversational-fast ($n = 10, M = 3.68, SD = 0.82$), formal-fast ($n = 10, M = 2.98, SD = 0.88$), conversational-slow ($n = 11, M = 2.96, SD = 1.07$), and formal-slow ($n = 10, M = 3.44, SD = 0.93$).

The two-way ANOVA revealed that the main effect of response speed was not statistically significant, $F(1,37) = 0.193, p = .663, \eta_p^2 = .005$. Similarly, the main effect of linguistic style was non-significant, $F(1,37) = 0.147, p = .704, \eta_p^2 = .004$. However, the interaction effect between response speed and linguistic style was $F(1,37) = 4.060, p = .051, \eta_p^2 = .099$. While this result sits at the margin of the conventional $.05$ significance level, it reached the threshold for significance at the $.10$ level. Given that it did not strictly meet the standard social science threshold of $p < .05$, definitive conclusions regarding the interaction should be drawn with caution. Nevertheless, the effect size ($\eta_p^2 = .099$) is considerably larger than those of the individual main effects and represents the largest effect size among the affective indicators analyzed. This implies that social presence may be more sensitive to the *combination* of response speed and linguistic style rather than their independent influences.

The interaction exhibited a clear crossover pattern. In the conversational condition, social presence was higher with fast responses (Group B) than with slow responses (Group D) ($3.68 > 2.96$). Conversely, in the formal condition, social presence was higher with slow responses (Group E) than with fast responses (Group C) ($3.44 > 2.98$). These findings suggest that rapid response speed does not universally enhance presence; rather, the experience of social presence is governed by the **congruence** between actual speed and the expectations established by the linguistic tone. This trend can be summarized as follows: while the conversational style fosters an expectation for interactional immediacy—leading to higher presence when met and significant declines when delayed—the formal style does not necessarily link latency to decreased presence, showing higher scores in the slow condition.

These results indicate that social presence is not merely a function of response speed or linguistic tone in isolation but is likely related to the learner's interpretive processing of the interactional context. Specifically, while a conversational style evokes expectations aligned with human-like dialogue norms, a formal style may establish a cognitive frame that permits more deliberative, reflective responses. As this study did not directly manipulate or measure these underlying psychological mechanisms, these causal interpretations are positioned as exploratory possibilities to be discussed in Chapter 7. Such interpretations provide a foundation for future research to incorporate additional factors, such as status indicators, turn-taking signals, and more granular linguistic variations.

Table 14. Cell Descriptive Statistics for Social Presence

Language Style	**Response Speed: Fast** **Mean(SD), N**	**Response Speed: Slow** **Mean(SD), N**
Conversational Tone	3.68 (0.82), n = 10	2.96 (1.07), n = 11
Formal Tone	2.98 (0.88), n = 10	3.44 (0.93), n = 10

Table 15. Two-Way ANOVA Results for Social Presence

Effect	**F(1,37)**	**p**	**Partial Eta Squared (ηp^2)**
Response Speed (Fast vs. Slow)	0.193	0.663	0.005
Language Style (Conversational vs. Formal)	0.147	0.704	0.004
Response Speed × Language Style	4.060	0.051	0.099

Table 16. Levene's Test for Social Presence

Test	**p**
Levene's Test for Homogeneity of Variances	0.781

Figure 5. Estimated Marginal Means of Social Presence

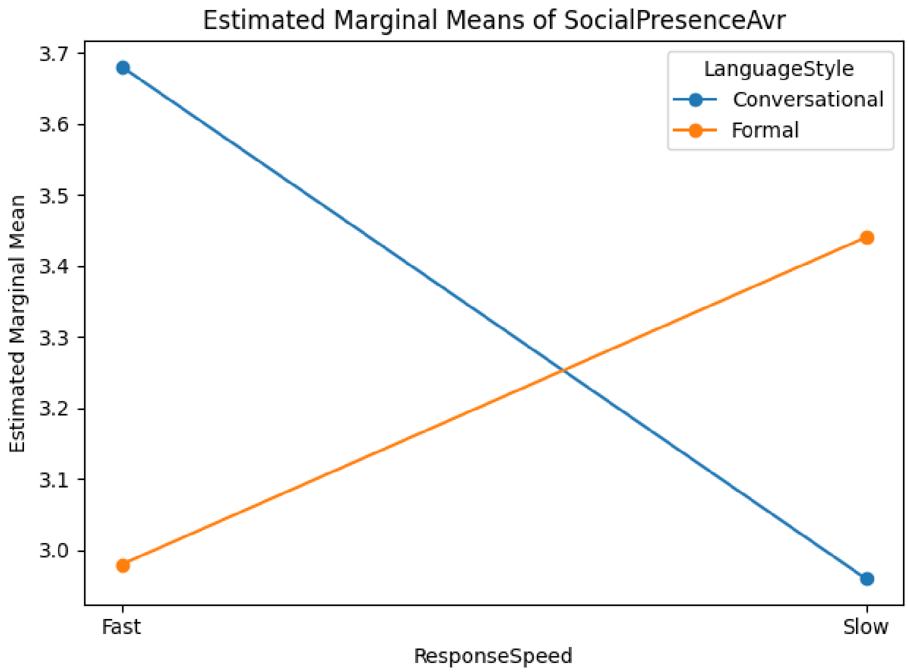


Figure 5 illustrates the estimated marginal means of social presence (SocialPresenceAvr) as a function of response speed and linguistic style. A distinct crossover interaction pattern is observed between the two variables. In the conversational condition, social presence reached its peak with fast responses ($M = 3.68$) and decreased significantly when latency was introduced ($M = 2.96$). Conversely, in the formal condition, social presence was higher with slow responses than with fast ones ($M = 3.44$ vs. 2.98). Thus, social presence does not follow a simple linear relationship where "faster is always better"; rather, the directional effect of response speed is reversed depending on the interactional context established by the linguistic style.

These findings partially deviate from the initial hypothesis, which predicted that fast responses would consistently enhance social presence. However, when reinterpreted through the lens of contextual congruence between linguistic and temporal cues, they reveal a consistent behavioral pattern. The conversational style creates an expectation of immediacy similar to everyday dialogue, which is satisfied when paired with fast responses. In this case, learners perceive the AI avatar as a social entity akin to a real conversation partner, thereby maximizing social presence. Conversely, when responses are delayed in the conversational condition, this expectation is frustrated, leading to a sharp decline in presence.

On the other hand, the formal condition exhibited a contrasting trend. Formal language carries social cues that emphasize expertise, prudence, and accuracy. In this context, a degree of latency may be interpreted as a "deliberative thinking process" before providing a response. The fact that the formal-slow condition yielded higher social presence than the formal-fast condition suggests that learners may have reinterpreted the delay not as a technical glitch, but as a sign of careful cognitive processing. This implies that the "professional image" of the formal persona provided a buffering effect, transforming response latency into a socially acceptable cue.

In summary, the core message of Figure 5 is that social presence is governed not by the independent effects of response speed or linguistic style, but by the integrated interaction persona formed by their combination. The conversational persona maximizes social presence when paired with fast responses, while the formal persona does so when combined with relatively slow responses. These results indicate that learners do not perceive linguistic tone and response rhythm as isolated elements but integrate them into a coherent social behavioral pattern. Therefore, this study demonstrates that simply maximizing response speed is not always the optimal strategy for AI avatar design. Instead, designing response rhythms that align with the established linguistic style provides a more sophisticated design principle for maximizing social presence. This serves as a primary empirical basis for the concept of "Social Cue Alignment" to be discussed in Chapter 7.

5-1-3 Sv n, v `x @MNU@ enq Bnf mhshud Kn`c 'Bnf mhshudKn`c @uq(

A 2×2 two-way analysis of variance (ANOVA) was conducted with cognitive load (**CognitiveLoadAvr**) as the dependent variable to examine the effects of response speed (fast vs. slow) and linguistic style (conversational vs. formal). Prior to the analysis, Levene's test for equality of error variances yielded a non-significant result ($p = .601$), confirming that the assumption of homogeneity of variance was satisfied (Table 19). The sample sizes for each condition were as follows: conversational-fast ($n = 10$), formal-fast ($n = 10$), conversational-slow ($n = 11$), and formal-slow ($n = 10$). The mean scores were 1.97 (Group B), 2.53 (Group C), 2.58 (Group D), and 2.40 (Group E), with standard deviations of 0.60, 0.91, 0.68, and 0.72, respectively (Table 17).

The two-way ANOVA results indicated that the main effect of response speed was not statistically significant—$F(1,37) = 1.077, p = .306, \eta_p^2 = .028$. Similarly, the main effect of linguistic style did not reach significance ($F(1,37) = 0.727, p = .399, \eta_p^2 = .019$), and no significant interaction effect was observed ($F(1,37) = 2.622, p = .114, \eta_p^2 = .066$). However, it is noteworthy that the effect size of the interaction was relatively larger than those of the individual main effects. This suggests that cognitive load may be determined by the specific combination of response speed and linguistic style rather than by either factor in isolation.

The estimated marginal means illustrated in Figure 6 reveal a crossover interaction pattern similar to the one observed for social presence. In the conversational condition, cognitive load was at its lowest when the response was fast ($B = 1.97$) but increased sharply when the response was delayed ($D = 2.58$). In contrast, within the formal condition, the difference in cognitive load between fast and slow responses ($C = 2.53$ vs. $E = 2.40$) was relatively minimal, suggesting that cognitive burden in this condition was less sensitive to variations in speed. In essence, while latency significantly exacerbated the burden in the conversational style, the same delay was perceived as relatively less burdensome in the formal style.

These results indicate that response latency is not merely a matter of increased waiting time. A conversational tone tends to establish an expectation that a response will be forthcoming immediately; when this expectation is violated, users may begin to doubt the system's functional integrity. For instance, prolonged silence may lead learners to speculate whether an error occurred or if their query was even transmitted—a deliberative process that acts as an additional cognitive burden. This likely explains why cognitive load was highest in the conversational-slow condition. Conversely, a formal linguistic style conveys a more deliberate and official impression, allowing learners to interpret delays as a sign of careful processing rather than a technical failure. Consequently, even with slower responses, the pattern did not show a sharp increase in cognitive load.

This interpretation is further supported by the qualitative feedback discussed in Section 6.5, where several participants reported that delayed responses felt like the system had "frozen" and expressed a need for status indicators. In summary, the findings regarding cognitive load demonstrate that the congruence between response speed and linguistic style is more critical than the independent effect of either variable. Specifically, for AI avatars utilizing a conversational persona, failing to provide rapid responses may inadvertently increase learner burden. This suggests that when designing AI-based educational systems, a friendly persona should be coupled with either reduced latency or supporting UX features, such as loading icons or "generating response" notifications, to mitigate cognitive uncertainty.

Table 17. 2 × 2 Cell Descriptive Statistics for Cognitive Load

Language Style	Response Speed: Fast	Response Speed: Slow
Conversational Tone	1.97 (0.60), n = 10	2.58 (0.68), n = 11
Formal Tone	2.53 (0.91), n = 10	2.40 (0.72), n = 10

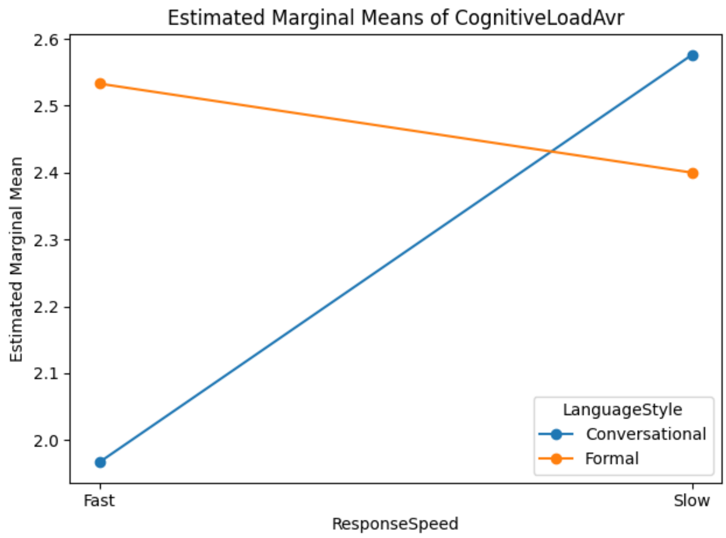
Table 18. Two-Way ANOVA Results for Cognitive Load

Effect	F(1,37)	p	Partial Eta Squared (ηp^2)
Response Speed (Fast vs. Slow)	1.077	0.306	0.028
Language Style (Conversational vs. Formal)	0.727	0.399	0.019
Response Speed × Language Style	2.622	0.114	0.066

Table 19. Levene's Test for Cognitive Load

Test	p
Levene's Test for Homogeneity of Variances	0.601

Figure 6. Estimated Marginal Means of Cognitive Load



6.3 Verification of Mediating and Moderated Mediating Effects (H2)

Hypothesis 2 (H2) posits that interaction characteristics (Response Speed × Linguistic Style) indirectly influence learning performance (ScoreDiff) through the mediators of flow (FlowAvr), social presence (SocialPresenceAvr), and cognitive load (CognitiveLoadAvr). To test this, a moderated mediation analysis was performed using PROCESS Macro (Model 8), and the results are summarized in Table 20 and Table 21.

First, regarding the pathway mediated by flow, the 95% bootstrap confidence intervals (CIs) for the conditional indirect effects included zero across all conditions, indicating a lack of statistical significance. Furthermore, the index of moderated mediation was not significant. This result is consistent with the findings in Section 6.2.2, where the interaction effect on flow was nearly zero ($F(1,37) = 0.000, p = .984$), suggesting that interaction characteristics do not influence achievement gains via flow.

Second, for the pathway mediated by social presence, the conditional indirect effects were likewise non-significant in all conditions. Although the two-way ANOVA in Section 6.2.3 revealed a marginal interaction effect at the social presence stage ($F(1,37) = 4.060, p = .051$), statistical evidence was insufficient to confirm that this effect sequentially mediated learning performance (ScoreDiff). In other words, while the data reflect a specific interaction pattern for social presence, they do not support its structural role as a mediator for achievement change.

Third, in the pathway mediated by cognitive load, neither the conditional indirect effects nor the index of moderated mediation reached statistical significance. This implies that the crossover pattern observed in cognitive load (as described in Section 6.2.4) did not translate into a statistically significant mediation of learning performance.

In summary, Hypothesis 2 (H2)—which predicted that the interaction between response speed and linguistic style would indirectly affect learning performance through flow, social presence, and cognitive load—was not supported. Rather than attempting to statistically finalize a mediation mechanism, it is more appropriate to engage in an exploratory discussion of potential pathways in Chapter 7 by synthesizing the quantitative patterns and qualitative feedback. Given the pilot nature of this study, these findings underscore the necessity for future research to re-examine the mediation structure with an expanded sample size.

6.4 Detailed Report of Quantitative Survey Findings

Distinct from the formal hypothesis testing presented in Section 6.2, this section provides a more granular examination of the "mean profiles" across each experimental condition. From an evaluative perspective, the primary role of this section is to ensure methodological transparency by reporting the substantive patterns inherent in the data, even in instances where statistical significance was not reached. By detailing these descriptive trends, this section offers a comprehensive view of the practical implications and nuanced learner experiences observed within the study's specific context.

Table 20. Moderated Mediation of Response Speed (PROCESS Model 8)

Predictor → Mediator → Outcome	RxSpeed → FlowAvr → SDiff	RxSpeed → SocAvr → SDiff	RxSpeed → CogLoad → SDiff
Conditional Indirect Effect (Conversational)	-0.1413	-0.2468	-0.0836
95% BootCI (LL, UL)	[-1.0089, 0.5023]	[-1.0308, 0.2773]	[-0.6376, 0.3985]
Conditional Indirect Effect (Formal)	-0.1488	0.1585	0.0183
95% BootCI (LL, UL)	[-0.7680, 0.3644]	[-0.2325, 0.7080]	[-0.2903, 0.3641]
Index of Moderated Mediation	-0.0075	0.4053	0.1019
95% BootCI (LL, UL)	[-0.8827, 0.9547]	[-0.4207, 1.4953]	[-0.5616, 0.8566]

Table 21. Moderated Mediation of Language Style (PROCESS Model 8)

Predictor → Mediator → Outcome	RxSpeed → FlowAvr → SDiff	RxSpeed → SocAvr → SDiff	RxSpeed → CogLoad → SDiff
Conditional Indirect Effect (Fast Response)	-0.1819	-0.2412	-0.0778
95% BootCI (LL, UL)	[-1.0597, 0.3840]	[-0.8392, 0.3072]	[-0.5818, 0.3803]
Conditional Indirect Effect (Slow Response)	-0.1894	0.1641	0.0241
95% BootCI (LL, UL)	[-0.8800, 0.3948]	[-0.1999, 0.9170]	[-0.2550, 0.3746]
Index of Moderated Mediation	-0.0075	0.4053	0.1019
95% BootCI (LL, UL)	[-0.8651, 0.9446]	[-0.4153, 1.5528]	[-0.5563, 0.8137]

5-3-0 Bnl o`qhrnmne L d`mr ax P%@@u`hk`ahkhsx 'Bnmsqnkur-Onnkdc Dwodqhl dms`k(

When reconfiguring the data in Table 3 based on the availability of the Q&A system, the control group reported mean scores of 3.29 for flow, 3.04 for social presence, and 2.73 for cognitive load. Among the experimental conditions, Group B exhibited the most positive profile, with flow at 4.08, social presence at 3.68, and cognitive load at 1.97. Descriptively, the provision of a Q&A system appears to shift the profile toward enhanced flow and satisfaction while simultaneously reducing cognitive load. It should be noted, however, that these observations represent "mean patterns," and their statistical significance must be interpreted independently based on the ANOVA results for each variable

5-3-1 Bnl o`qhrnmne L d`mr ax Khmftrshb Rsxkd 'Bnmudqr`shnm`kur-Enql`k(

Conditions utilizing a conversational style (B, D) reported relatively higher flow scores ($4.08\ and\ 3.91$, respectively) compared to those using a formal style (C, E; $3.86\ and\ 3.68$, respectively). However, the patterns for social presence were more complex: while the conversational style yielded very high presence when paired with fast responses (Group B: 3.68), this value decreased significantly when paired with slow responses (Group D: 2.96). A similar trend was observed for cognitive load, which was at its lowest in the conversational-fast condition (Group B: 1.97) but increased in the conversational-slow condition (Group D: 2.58). These results reinforce the notion that linguistic style should not be viewed as an isolated factor; rather, its impact depends on how the chosen style sets the learner's expectations for response speed.

5-3-2 Bnl o`qhrnmne L d`mr ax Qdronmrd Roddc 'E`rs ur-Rknv(

In the fast response conditions (B, C), social presence scores diverged sharply depending on the linguistic style ($3.68\ vs.\ 2.98$). A similar divergence was observed in the slow response conditions (D, E), where social presence again varied significantly by style (2.96 vs. 3.44). This recurring pattern suggests that response speed possesses limited standalone explanatory power. Instead, its influence on the user experience is predominantly shaped through its interaction with the linguistic style.

5-3-3 Cdbntokhmf Kd`qmhmf Odqenql`mbd `mc TrdqDwodqhdmbd9-Od`j Odqenql`mbd-ur--Od`j Dwodqhdmbd-

A critical finding emerges from the comparison between objective achievement and subjective experience. Regarding learning performance (ScoreDiff), Group E demonstrated the highest mean gain (M₩ =₩ 1.50). Conversely, in terms of user experience dimensions—including flow, social presence, cognitive load, satisfaction, and subjective achievement—Group B received the most favorable overall evaluations. This indicates a clear dichotomy between the condition that yielded the greatest test score improvement and the condition that learners perceived as the most comfortable and engaging.

These results suggest that an enjoyable and low-burden learning experience does not necessarily translate into the highest level of academic achievement. Rather than viewing this as a limitation, it can be interpreted as a multi-objective design challenge for AI tutors. Specifically, it implies that the AI's persona—including linguistic style and response latency—should be strategically configured based on whether the primary educational goal is to maximize immediate objective achievement or to prioritize the overall quality of the learner's experience.

6.5 Qualitative Analysis of Survey Responses

This section categorizes open-ended responses into key themes to provide contextual evidence for the patterns observed in the quantitative results—specifically the decline in social presence and the increase in cognitive burden within the conversational-latency conditions. These qualitative data are positioned not as a substitute for statistical significance, but as complementary material that elucidates the underlying reasons for the observed behavioral and experiential trends.

5-4-0 Hlonqs`mbd ne Qdronmrd K`sdmbx `mc Rs`str Uhrhahkhsx

The most frequently recurring feedback concerned the discomfort caused by the lack of visibility during waiting periods after a query. When an immediate response was not forthcoming, participants often misinterpreted the delay as a system error or "freeze" rather than a cognitive processing phase for the avatar. Multiple respondents noted, "I thought it was an error when the avatar didn't answer immediately; it would be better if the 'thinking' status were intuitively displayed," and "There should be some indicator during the waiting time to show the system hasn't crashed."

These reactions align with the interaction patterns for social presence identified in Section 6.2.3, particularly the decrease in presence within the conversational-slow group. Because a conversational style sets an expectation for human-like interaction, users naturally anticipate immediate feedback. When this expectation is violated by latency, users may perceive a break in the interaction or assume a system malfunction, thereby diminishing their sense of social presence. While a clear status indicator (e.g., "Generating response...") might mitigate this decline, this remains a design-level proposition for discussion in Chapter 7, as the current study did not manipulate status visibility.

5-4-1 Unhbd Pt`khsx `mc Odqbdoshmmne Sdbgmhb`k Fkhsbgdr

Qualitative feedback also highlighted that inaccuracies in the avatar's pronunciation were often conflated with system errors. One participant explicitly stated, "The avatar's pronunciation was imprecise in some parts, which made me think a technical glitch had occurred." This suggests that unnatural prosody or phonetic lack of clarity can lead users to interpret auditory quality issues as fundamental system failures.

Such responses underscore the significant impact of text-to-speech (TTS) quality, including pitch and intonation, on user experience. When a delayed response is compounded by awkward vocal delivery, users are more likely to conclude that the "system is not functioning correctly." This can erode trust in the AI tutor and increase cognitive load as learners expend additional mental effort to interpret the ambiguous situation. Consequently, phonetic clarity is a critical factor influencing both social presence and cognitive load.

5-4-2 Bnmsqhatshnm ne Mnm,udqa`k Dwoqdrrhnmr sn Rnbh`k Oqdrdmbd

Several participants noted that unnatural movements or appearances hindered their sense of presence. Comments included observations that "the avatar's appearance felt unnatural" and that "adding gestures like eye movements or hand motions would make it feel more human-like." This indicates that learners perceive visual cues as being just as significant as linguistic style and response speed.

These findings clarify that social presence is not determined solely by verbal or temporal cues. Even when an optimal combination of style and speed is achieved, the overall sense of presence may be constrained if non-verbal cues—such as facial expressions, gaze, and gestures—lack naturalness. Future research should control for the quality of these non-verbal expressions or incorporate them as independent variables to distinguish their effects from linguistic and temporal factors.

5-4-3 Cdl`mc enq Rtashskdr9 Rs`ahkhsx `mc @bbdrrhahkhsx

A clear demand for simultaneous subtitle support emerged from the qualitative data. Participants suggested that "it would be helpful if the answers appeared as subtitles" and that "text support would make the audio easier to understand." This request appears to be driven by a need to reduce the cognitive friction experienced during the learning process.

Subtitles provide a redundant visual channel that allows learners to verify information that might be difficult to grasp through audio alone, especially when pronunciation or intonation is unnatural. Furthermore, in cases of response latency, the appearance of text or a subtitle placeholder can serve as a signal that "an explanation is forthcoming," allowing the learner to wait more patiently. Thus, subtitle support is a design element that can reduce cognitive burden and foster a more seamless interaction.

5-4-4 Rtl l `qx ne Pt`kks`shud Ehmchmf r

In reporting these qualitative findings, this section maintains a descriptive focus to avoid over-interpreting causality. The feedback is presented as a record of observed participant perceptions—for example, noting that "some responses indicated that latency was mistaken for system failure" or that "users expressed a need for status indicators and subtitle support." Similarly, observations that "non-verbal enhancements could improve social presence" are reported as thematic presence within the data. This approach ensures that the qualitative evidence remains a grounded supplement to the quantitative results without exceeding the interpretative boundaries appropriate for this chapter.

6.6 Synthesis of Qualitative Feedback and Integrated Interpretation

This section synthesizes the user feedback collected from open-ended responses and integrates it with the quantitative results to provide a holistic interpretation of the learner experience. The objective is not to assert definitive causality but to elucidate the subjective experiences of learners during the experimental process and derive substantive design implications.

The qualitative analysis revealed several recurring themes. Most notably, when responses were not instantaneous, participants tended to perceive the latency as a system error or malfunction rather than a deliberative "thinking" phase of the AI avatar. Respondents frequently noted that significant delays made the system feel "frozen" and left them "uncertain whether their queries had been successfully transmitted." Consequently, there were explicit demands for enhanced status visibility, such as loading animations or "generating response" indicators, to mitigate the anxiety associated with waiting. Furthermore, participants highlighted the necessity of simultaneous subtitle support and more natural non-verbal cues (e.g., eye blinking, head movements, and gestures) to bolster social presence and anthropomorphism.

These qualitative insights provide a critical context for understanding the interaction patterns observed in the quantitative analysis, particularly the marginal interaction effect in social presence ($p = .051$). In the quantitative data, the conversational style fostered high presence when paired with fast responses (Group B, $M = 3.68$) but suffered a sharp decline when latency was introduced (Group D, $M = 2.96$). Conversely, the formal style appeared more congruent with slower response times (Group E, $M = 3.44$). Cognitive load patterns mirrored these trends, showing the greatest increase in the conversational-slow condition.

Triangulating these findings suggests that without clear status indicators, learners often misinterpret response latency as technical instability. This perception is particularly acute in conversational conditions, where the linguistic style sets an expectancy for immediate reciprocity. When these expectations are violated by silence and a lack of visual cues, social presence erodes and cognitive burden increases due to the effort required to interpret the system's "silence." In contrast, a formal style may evoke a "professional lecture" frame that allows for more tolerance toward deliberative pauses.

In summary, the quality of the learner experience is not determined by linguistic style or response speed in isolation. Instead, it is a product of the contextual congruence between the AI's persona and supporting UX elements such as status indicators, subtitles, and non-verbal expressions. These findings underscore that the overall interactional context is a fundamental determinant of social presence and cognitive load in AI-driven educational environments.

7. Conclusions and Discussion

This chapter synthesizes the quantitative and qualitative findings presented in Chapter 6 to draw conclusions regarding the research questions and hypotheses, while discussing their broader theoretical and design implications. Acknowledging that the primary hypotheses were not supported at the 5% significance level, this discussion avoids overstating statistical significance. Instead, it integrates descriptive data patterns, effect sizes, and qualitative participant feedback to derive limited but substantive insights for the design of AI avatar-based online lecture interactions.

7.1 Reinterpreting the Efficacy of Real-time Q&A: A Focus on Achievement Stability

In this study, the presence of a real-time Q&A function did not yield a statistically significant mean difference in learning performance. However, descriptive statistics revealed a noteworthy trend: the mean achievement gain for the experimental groups ($M = 0.9756$) was higher than that of the control group ($M = 0.5455$). More critically, a stark contrast was observed in the dispersion of scores. While the control group exhibited high volatility ($SD = 5.1257$), the experimental groups demonstrated a markedly smaller standard deviation ($SD = 2.1271$). These results suggest that the real-time Q&A function may serve a vital role in reducing the variability of learning achievement, even if its impact on shifting the overall mean is not statistically definitive.

A detailed comparison of pre- and post-test score distributions further supports this interpretation. In the control group, scores remained widely dispersed across both assessments. In contrast, despite varying initial distributions, the experimental groups showed a clear convergence toward the upper bound (near-perfect scores) in the post-test. This shift indicates that real-time Q&A availability likely functioned as a mechanism to mitigate achievement gaps among learners.

While it is impossible to entirely rule out the possibility that these effects stemmed from idiosyncratic factors such as the participants' short-term memory, this study posits that the process of resolving ambiguities in real-time allowed learners to clarify and solidify core concepts. Thus, the provision of real-time Q&A acts as a stabilizing safeguard that prevents drastic achievement declines and raises the "achievement floor" for all learners. From this perspective, the efficacy of the Q&A system should be evaluated not just as a "mean-shifting" effect, but as a variance-reduction effect that enhances the reliability and consistency of educational outcomes.

This interpretation is reinforced by qualitative feedback. Participants noted that "being able to ask questions made it feel like a live class, which improved focus" and "immediate clarification of unknown parts helped with immersion." These sentiments align with the observed reduction in score dispersion. Consequently, incorporating real-time Q&A into AI avatar-based systems should be viewed as a structural device for ensuring the consistency of performance.

Conventional online video lectures often suffer from the limitation that learners cannot immediately resolve misunderstandings, leading to a cumulative "comprehension deficit." Furthermore, the act of pausing a lecture to search for information on external devices inevitably disrupts the learning flow. To transcend mere functional integration, this study suggests that the Q&A feature should be embedded as a "procedural regularity." For example, systems could include explicit prompts such as "Any questions?" at designated intervals or provide specific "question breaks" similar to offline settings. Such design interventions can lower psychological barriers and encourage active interaction by signaling to the learner that questioning is an integral part of the curriculum rather than an interruption.

7.2 Potential Decoupling of Learning Performance and Experience Quality

The statistical analysis of this study suggests a potential decoupling between objective learning performance ($PostTest - PreTest$) and affective experience indices (flow, social presence, cognitive load, satisfaction, and perceived achievement). Specifically, while the most significant improvement in learning achievement was observed in the formal-slow condition (Group E), the quality of user experience—represented by flow, social presence, and satisfaction—was generally higher in the conversational-fast condition (Group B). This indicates that the condition yielding the highest academic gain does not necessarily coincide with the condition perceived as most favorable by the learner.

These findings imply that a single "optimal strategy" may not exist in the design of AI avatar-led online lectures. Strategies for maximizing learning outcomes and those for enhancing user experience may diverge. Therefore, rather than limiting design goals to a single metric such as test scores, it is theoretically necessary to adopt a multi-objective optimization perspective that simultaneously accounts for both academic achievement and the quality of the learning experience. Future AI educational systems must strategically navigate the balance between the "interface for maximum achievement" and the "interface for maximum engagement."

7.3 Context-Dependent Optimality of Style-Speed Dyads

A critical pattern observed in the analyses of social presence and cognitive load is that the combination of factors—rather than their individual effects—governs the user experience. A marginal interaction effect was observed in social presence ($p = .051$); while the conversational-fast combination (Group B) fostered a strong sense of presence, the conversational-slow combination (Group D) resulted in a significant decline. Conversely, in formal conditions, social presence remained relatively stable even when response latency was introduced.

A similar crossover was observed in cognitive load. In conversational conditions, increased latency led to a sharp rise in cognitive burden, whereas in formal conditions, identical delays resulted in a more tempered increase. These results refute the simplistic rule that "faster is always better." Instead, they highlight that adjusting response latency to match the expectancy frame established by the linguistic style is a more crucial design factor.

As Large Language Models (LLMs) become more integrated into educational settings, users have become increasingly accustomed to the "inference time" required for complex processing. Thus, response latency is not inherently negative. What matters is not the absolute duration of the delay, but its semantic congruence with the interaction persona. These findings suggest that conversational personas should be paired with minimal latency, while formal personas may tolerate longer delays without compromising presence or performance. This necessitates an adaptive latency strategy tailored to the avatar's linguistic persona.

7.4 Consistency of Conversational Effects on Flow and Theoretical Alignment

The analysis of flow scores revealed no significant interaction effect between response speed and linguistic style ($p = .984$), nor were there significant main effects. However, descriptive data indicated that across all speed conditions, participants generally reported higher flow when interacting with a conversational avatar. This suggests that flow may be more sensitive to linguistic style than to response speed. Utilizing a conversational style in AI-driven lectures may therefore serve as a robust emotional stabilizer that enhances learner engagement.

These results align with Mayer's (2021) Personalization Principle, which posits that informal, conversational language reduces extraneous cognitive load and strengthens social bonding and motivation. The observed trends in flow in this study support this theoretical framework. From a design standpoint, it is reasonable to adopt a conversational persona as the default for lectures where learner flow is the primary objective.

7.5 Persona Differentiation Based on Learner Expertise: Possibilities and Limits

Analysis by prior learning experience suggests that the optimal interaction condition may vary according to the learner's level of expertise. For novices (those without prior experience in social research methodology), the conversational-fast condition (B) yielded the highest mean gains. Conversely, for experienced learners, the formal-fast (C) and formal-slow (E) conditions were associated with higher achievement. This implies that a friendly, immediate-feedback system is effective for those encountering a subject for the first time, while a formal, deliberative persona may be more beneficial for those engaged in review or advanced study.

However, due to the exceedingly small sample sizes in these sub-categories ($n = 3\ to\ 7$), these results cannot be presented as definitive conclusions. This study characterizes these findings as "observed tendencies," underscoring the need for future research to systematically validate these hypotheses with larger samples and randomized assignment based on expertise levels.

7.6 Complementary UX Design Elements

Qualitative feedback highlighted several key areas for improvement, including the addition of non-verbal cues (eye movements, gestures, facial expressions), the provision of subtitles, and the necessity of status indicators during response delays. These insights are particularly vital for interpreting cognitive load patterns. The increase in cognitive load may stem less from the latency itself and more from the cognitive uncertainty caused by the absence of a "system active" signal. Much like a user attempting to refresh a frozen webpage, a lack of feedback from an AI avatar induces psychological stress.

To mitigate this, auxiliary UX elements should be considered. Displaying "thinking" expressions or loading indicators can provide a visual signal of active processing, while optional subtitles can serve as a psychological safeguard. Subtitles not only support auditory processing but also reduce misunderstandings regarding pronunciation and intonation while signaling that information is forthcoming. Although not manipulated as experimental variables in this study, these UX signals warrant further empirical investigation in future research as critical tools for managing cognitive load and social presence.

7.7 Conclusion

This study experimentally investigated the impact of real-time Q&A availability—moderated by response speed and linguistic style—on learning performance and experience within an AI avatar-based online lecture environment. Although the hypotheses derived from prior literature did not reach statistical significance, the study provides meaningful evidence that the Q&A-enabled groups outperformed the control group in mean achievement, and that conversational styles consistently fostered higher flow compared to formal styles. Based on the analysis of mean gain, effect sizes, and qualitative interviews, the findings are summarized into the following six points:

- Achievement Stability

  Groups utilizing real-time Q&A demonstrated smaller variance in score improvements compared to the control group. This suggests that the system functions as a stabilizing mechanism that reduces the volatility of learning outcomes and raises the achievement floor.

- Decoupling of Performance and Experience

  The condition yielding the highest academic gain did not always align with the condition providing the highest flow and social presence. This implies that academic achievement and learner experience may not be maximized simultaneously under a single condition, necessitating a balanced, multi-objective design approach.

- Contextual Efficacy of Response Speed

  Rapid response speed is not a universal desideratum. Its efficacy is contingent upon linguistic style; while fast responses were beneficial in conversational contexts, their impact was limited in formal settings, suggesting that the "style-speed" combination dictates the quality of interaction.

- Linguistic Style as a Robust Driver of Flow

  Across all speed conditions, conversational language led to higher average flow scores than formal language. This indicates that linguistic style exerts an independent and significant influence on the affective dimensions of learning.

- Expertise-Based Persona Tailoring

  For novice learners with no prior knowledge of the subject, the conversational-fast combination was most effective. In contrast, experienced learners showed greater gains in formal conditions. This suggests that the optimal interaction persona should be tailored to the learner's expertise level.

- UX Augmentation for Cognitive Support

  Qualitative feedback underscored the necessity of auxiliary UX elements, such as subtitles, status indicators, and non-verbal cues (e.g., gaze and gestures). These elements are critical for mitigating cognitive load and reinforcing social presence in AI-driven environments.

Regarding limitations, the recruitment of $N = 52$ participants resulted in relatively small cell sizes ($10 - 11$ per group), which lacked sufficient statistical power for confirmatory validation. However, this study was explicitly designed as an exploratory pilot study aimed at identifying directional patterns and effect sizes. The structural consistency observed in the mean patterns and effect sizes suggests that statistical significance could be attained with a larger sample. Therefore, these results should serve as a foundational basis for sample size estimation and variable optimization in future large-scale experiments.

In conclusion, this research experimentally demonstrates that micro-design variables, such as response speed and linguistic style, can exert divergent influences on learning outcomes and experience quality. As a point of departure for future research—including expanded sample sizes and the verification of complex mediating effects—this study contributes to building a robust framework for interactive, AI avatar-based online education.

8. References

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Briz-Ponce, L., Pereira, A., Carvalho, L., Juanes-Méndez, J. A., & García-Peñalvo, F. J. (2017). Learning with mobile technologies – Students' behavior. *Computers in Human Behavior*, *72*, 612–620. https://doi.org/10.1016/j.chb.2016.05.027

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (1.

HarperPerennial ed). HarperPerennial.

Georgiou, Y., & Kyza, E. A. (2018). Relations between student motivation, immersion and learning outcomes in location-based augmented reality settings. *Computers in Human Behavior*, *89*, 173–181. https://doi.org/10.1016/j.chb.2018.08.011

Herbert, C., & Dołżycka, J. D. (2024). Teaching online with an artificial pedagogical agent as a teacher and visual avatars for self-other representation of the learners. Effects on the learning performance and the perception and satisfaction of the learners with online learning: Previous and new findings. *Frontiers in Education*, *9*, 1416033. https://doi.org/10.3389/feduc.2024.1416033

Islam, M. Z., & Wang, G. (2025). Avatars in the educational metaverse. *Visual Computing for Industry, Biomedicine, and Art*, *8*(1), 15. https://doi.org/10.1186/s42492-025-00196-9

Kodani, N., Uchida, T., Kameo, N., Sakai, K., Funayama, T., Minato, T., Kikuchi, A., & Ishiguro, H. (2025). Android avatar improves educational effects by embodied anthropomorphization. *Frontiers in Robotics and AI*, *11*, 1469626. https://doi.org/10.3389/frobt.2024.1469626

Mayer, R. E. (2021). *Multimedia learning* (Third edition). Cambridge University Press.

https://doi.org/10.1017/9781316941355

Nah, F. F.-H. (2004). A study on tolerable waiting time: How long are Web users willing to wait? *Behaviour & Information Technology*, *23*(3), 153–163. https://doi.org/10.1080/01449290410001669914

Pang, H., Tang, S., Han, J. Y., & Fung, F. M. (2023). Exploring the Use of an Avatar-Based Online Platform to Facilitate Social Interaction in Laboratory Sessions. *Journal of Chemical Education*, *100*(10), 3832–3840. https://doi.org/10.1021/acs.jchemed.3c00093

Response Time Limits: Article by Jakob Nielsen. (n.d.). Nielsen Norman Group. Retrieved November 4, 2025, from https://www.nngroup.com/articles/response-times-3-important-limits/

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In Advances in Experimental Social Psychology (Vol. 19, pp. 123–205). Elsevier. https://doi.org/10.1016/S0065-2601(08)60214-2

Sherry, J. L. (2004). Flow and Media Enjoyment. *Communication Theory*, *14*(4), 328–347. https://doi.org/10.1111/j.1468-2885.2004.tb00318.x

Sinlapanuntakul, P., & Zachry, M. (2025). Perception in Pixels: Effects of Avatar

Representation in Video-Mediated Collaborative Interactions. *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work*, 1–16. https://doi.org/10.1145/3729176.3729183

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Ukenova, A., Bekmanova, G., Zaki, N., Kikimbayev, M., & Altaibek, M. (2025). Assessment and Improvement of Avatar-Based Learning System: From Linguistic Structure Alignment to Sentiment-Driven Expressions. *Sensors*, *25*(6), 1921. https://doi.org/10.3390/s25061921

Xia, T., Tan, M., & Guo, J. (2024). Avatars improve the quality of online learning: The role of attention and positive emotions. *Education and Information Technologies*, *29*(18), 24203–24222. https://doi.org/10.1007/s10639-024-12790-0