

Traditional vs. Socratic: How Chatbot Interaction Strategies Shape Self-Directed Learning?

Thiago J. Lopes¹, Yeshuá Siqueira¹, Jairo F. de Souza^{1,2}, Marcelo Machado^{1,2}

¹Laboratory of Innovative Applications in Computing (LApIC)
Federal University of Juiz de Fora (UFJF)
Juiz de Fora – MG – Brazil

²Department of Computer Science
Federal University of Juiz de Fora (UFJF)
Juiz de Fora – MG – Brazil

{thiago.jose, yeshua.siqueira}@estudante.ufjf.br
{jairo.souza, marcelo.machado}@ufjf.br

Abstract. *AI-based chatbots are increasingly used to support self-directed learning, yet little is known about how their interaction strategies shape learning. Grounded in the theory of Desirable Difficulties, we conducted a mixed-methods study investigating the effects of Traditional and Socratic interactions on undergraduate students learning Object-Oriented Programming. We analyzed immediate learning, knowledge retention, information behavior, and user experience. Exploratory findings suggest a trade-off between immediate performance and retention stability: Traditional interactions favored higher performance and greater efficiency, whereas Socratic interactions promoted more exploratory behavior and more stable knowledge retention.*

1. Introduction

Access to information has changed over the past decades. Search engines have long supported self-directed learning by enabling individuals to seek information and construct knowledge about unfamiliar topics. More recently, Generative Artificial Intelligence (GenAI) has further transformed this landscape [Hoppe et al., 2026], where AI-based conversational systems are increasingly used to explain concepts, answer questions, and support learning through natural dialogue.

Despite their educational potential, AI-based chatbots also introduce new challenges. Their ability to provide immediate, well-structured answers can reduce the cognitive effort required during learning, encouraging passive information consumption and creating a false sense of mastery [Gerlich, 2025, Lee et al., 2025]. This concern is consistent with research on cognitive offloading, which suggests that delegating cognitive processes to external systems can weaken memory encoding and long-term retention [Barcaui, 2025, Risko and Gilbert, 2016, Sparrow et al., 2011]. Consequently, an important question emerges: how should conversational AI interact with learners to support durable learning rather than simply deliver correct answers?

One promising perspective comes from the theory of Desirable Difficulties [Bjork et al., 2011, Bjork, 1994] which argues that introducing manageable cognitive challenges during learning can improve long-term retention, even when it reduces immediate performance. Rather than minimizing effort, effective instructional strategies should encourage

learners to actively retrieve, organize, and connect knowledge. Within conversational AI, this suggests that the educational effectiveness of a chatbot may depend not only on the quality of its responses but also on how those responses are managed.

A pedagogical approach aligned with this perspective is the Socratic method, in which instructors stimulate reasoning through guided questioning instead of simply providing direct answers [Paul, 1993, Reznitskaya and Gregory, 2009]. Recent studies indicate that Socratic AI systems can promote critical thinking and cognitive engagement [Fakour and Imani, 2025, Favero et al., 2024, Lee et al., 2026]. However, evidence regarding their impact on measurable learning outcomes remains limited. Moreover, existing work has primarily focused on learning performance or user engagement in isolation, leaving open the question of how different chatbot interaction strategies shape both learning outcomes and the information behaviors through which learning occurs.

Motivated by these gaps, this paper investigates how two distinct chatbot interaction strategies influence self-directed learning. We conducted a mixed-methods study involving undergraduate students learning Object-Oriented Programming (OOP) concepts. Participants interacted with either a Traditional chatbot that provided direct answers or a Socratic chatbot designed to promote reflective reasoning through guided questioning. We compared the two strategies considering immediate learning, delayed knowledge retention, interaction behavior, and user experience.¹ Our results provide evidence that chatbot interaction strategies influence not only learning outcomes but also the ways learners interact with AI during the learning process, highlighting interaction design as an important dimension of AI-supported learning environments.

The rest of this paper is structured as follows: Section 2 reviews the theoretical foundations of *S&CAL*, Desirable Difficulties, and Socratic pedagogy. Section 3 highlights the related works. Section 4 describes the experimental design. Section 5 presents and discusses the findings. Section 6 concludes the paper.

2. Background

This section presents the theoretical foundations that support our study: Search and Chat as Learning, the Socratic teaching method, and the theory of desirable difficulties.

2.1. Search and Chat as Learning

Research on learning through information systems has traditionally been framed within the Search as Learning (SAL) paradigm, which recognizes that the information-seeking process is not merely an information retrieval task but a learning process, in which effective search enhances learning outcomes, and learning equips individuals with the skills for efficient searching [Machado et al., 2020]. In this context, learning-oriented search goes beyond simple fact-finding, involving iterative processes such as sense-making, source evaluation, and the integration of new information into existing knowledge structures.

Recent work extends the SAL paradigm by treating conversational interaction as a learning process in its own right. Under perspectives such as *Chat as Learning* (CaL) [Yu and Liu, 2025] and *Chat-and-Search* [Tibau et al., 2024], learning emerges through iterative dialogue, where each conversational exchange helps learners refine their understanding while allowing AI systems to adapt subsequent responses. Compared with traditional

¹ All experiment data are available at: <https://github.com/lapic-ufjf/socratic-ai/>

search, conversational systems therefore play a more active role in shaping the learning process, although this also introduces challenges related to information reliability and overreliance on AI-generated answers.

At the same time, research has explored how conversational AI can be combined with traditional search systems. The *Search+Chat* paradigm integrates web search with embedded GenAI components to support learning-oriented tasks [Yang et al., 2025]. The findings suggest that users rely on chat for quick explanations and clarification, while still using search engines to verify information and explore topics more deeply, highlighting the complementary nature of these tools.

Taken together, these developments point to a broader paradigm, which we call *Search and Chat as Learning* (S&CAL): an evolution of SAL encompassing learning mediated by search engines and conversational AI, used independently, sequentially, or in combination within the same learning process.

2.2. Desirable Difficulties

The concept of desirable difficulties provides a theoretical foundation for understanding why effortful learning can lead to better outcomes. Introduced by Bjork [1994] and further developed by Bjork et al. [2011], it posits that learning conditions that require greater cognitive effort, when appropriately calibrated, enhance long-term retention and transfer.

Empirical research supports several forms of desirable difficulties, such as retrieval practice, spaced repetition, and generation tasks, which require learners to actively reconstruct knowledge and, despite increasing perceived difficulty, strengthen memory representations and transfer [Cepeda et al., 2006, Roediger and Karpicke, 2006].

Within S&CAL environments, this framework highlights a key trade-off between efficiency and learning depth. Systems that prioritize immediate answers may reduce cognitive effort and create a sense of understanding without promoting durable learning. In contrast, interaction designs that introduce productive challenges, such as prompting learners to think, explain, or predict, can foster more meaningful engagement.

In this regard, Socratic interaction can be seen as a natural instantiation of desirable difficulties. By replacing direct answers with guided questioning, Socratic AI-based chatbots should encourage learners to actively process information and construct their own understanding. This connection provides a principled basis for comparing Socratic and traditional chatbot designs with respect to their impact on learning.

2.3. The Socratic Teaching Method

The Socratic method is a dialogical approach to teaching that relies on structured questioning to stimulate reasoning and uncover assumptions. Rather than providing direct answers, the instructor guides learners through a process of inquiry in which understanding is gradually constructed [Boghossian, 2006, Paul and Elder, 2019].

A key feature of this approach is its emphasis on active cognitive engagement. Learners are encouraged to articulate their reasoning, examine inconsistencies, and refine their ideas through dialogue. This process aligns with evidence from the learning sciences showing that deeper learning occurs when learners actively construct knowledge rather than passively receive it [Chi, 2009]. In practice, Socratic questioning promotes critical thinking, argumentation, and metacognitive awareness [Reznitskaya and Gregory, 2009].

In the context of S&CAL, the Socratic method offers a distinct interaction model compared to both traditional search and answer-oriented chat systems. While conventional systems tend to optimize for efficiency by delivering immediate responses, a Socratic approach introduces guided inquiry as part of the interaction itself. This shift positions questioning not as a limitation, but as a mechanism for supporting deeper engagement and knowledge construction. As such, Socratic strategies can be operationalized in AI-based systems to scaffold reasoning while preserving learner autonomy.

3. Related Works

The rapid adoption of GenAI in education has led to a growing body of research investigating its impact on learning, behavior, and outcomes. Existing work provides three complementary perspectives on this problem: (i) the use of AI chatbots as learning tools, (ii) concerns regarding passive learning and cognitive offloading, and (iii) the design of pedagogically grounded AI systems, particularly those inspired by the Socratic method.

A first line of research investigates AI chatbots as learning tools. Conversational agents powered by large language models have been applied to support a wide range of learning activities, including tutoring, question answering, and personalized feedback [Duong et al., 2024, Kasneci et al., 2023, Ogunleye et al., 2024]. Among these studies, Blasco and Charisi [Blasco and Charisi, 2024] compared different AI interaction strategies, including direct answers, step-by-step explanations, and Socratic questioning. While Socratic interactions increased learner engagement and interaction frequency, they did not significantly improve learning outcomes or knowledge retention compared with answer-oriented systems. These findings suggest that interaction strategy alone is insufficient to guarantee deeper learning and highlight the need to better understand how different interaction styles influence the learning process.

A second research direction focuses on the potential drawbacks of AI-assisted learning. Studies on cognitive offloading argue that learners increasingly rely on AI systems to perform cognitive tasks that would otherwise contribute to learning [Risko and Gilbert, 2016, Sparrow et al., 2011]. Supporting this concern, a randomized controlled trial found that students using ChatGPT as a study aid performed significantly worse on a delayed retention test than those using traditional study methods [Barcaui, 2025]. This effect was attributed to reduced cognitive effort during learning, weakening memory encoding and consolidation. These findings are consistent with the theory of desirable difficulties [Bjork et al., 2011], suggesting that the immediacy of AI-generated answers may inadvertently reduce the effort required for durable learning. Consequently, an important open question is whether alternative interaction strategies can mitigate these effects while preserving the benefits of conversational AI.

Motivated by these concerns, a third line of research investigates conversational systems designed to actively promote learning rather than simply deliver answers. In particular, several studies have explored Socratic interaction strategies that encourage learners to construct knowledge through guided questioning [Degen, 2025, Fakour and Imani, 2025, Favero et al., 2024, Lee et al., 2026]. Across different educational contexts, these systems have consistently been associated with greater learner engagement, reflection, critical thinking, and metacognitive awareness. Nevertheless, evidence regarding their impact on measurable learning outcomes remains inconclusive. Most existing studies emphasize immediate performance, cognitive engagement, or perceived learning experi-

ence, while comparatively little attention has been given to delayed knowledge retention.

Taken together, these studies indicate that chatbot interaction design plays an important role in AI-supported learning. However, important gaps remain. Existing work has largely examined learning outcomes, cognitive engagement, or critical thinking in isolation, making it difficult to understand how different interaction strategies jointly affect learning performance and learner behavior throughout the learning process. Moreover, relatively few studies compare alternative AI interaction strategies while simultaneously evaluating delayed knowledge retention.

To address these gaps, this work compares two contrasting chatbot interaction strategies: a Traditional answer-oriented chatbot and a Socratic chatbot designed to scaffold learner reasoning through guided questioning. Adopting a mixed-methods perspective within the S&CAL framework, we jointly examine immediate learning, delayed knowledge retention, information behavior, and user experience. By combining quantitative learning measures with qualitative analyses of learner interactions, our study provides a broader understanding of how chatbot interaction strategies shape both learning outcomes and the information behaviors through which learning unfolds.

4. Study Design

This section describes the pilot study, the chatbot configurations, the participant procedures, and the evaluation measures.

4.1. Pilot Study

Prior to the main experiment, we conducted a pilot study involving 17 volunteer participants who were not involved in the final study. The pilot study used music theory as the learning domain and aimed to evaluate the experimental platform, collect preliminary user experiences, and refine the Socratic instruction (prompt).

Analysis of interaction logs and participant feedback revealed that early versions of the Socratic prompt occasionally produced excessive questioning, making conversations feel unnatural. Consequently, the prompt was redesigned to better balance guided questioning and direct explanations while preserving active reasoning. Participants perceived the music theory topics as having limited personal relevance, suggesting lower engagement than in authentic learning scenarios. Thus, we designed the main study to adopt OOP concepts, which are directly related to participants' academic activities. The pilot also revealed potential recall effects from identical immediate and delayed assessments, leading us to subtly rephrase the retention-test questions, changing only their wording while preserving conceptual equivalence.

4.2. Experimental Setup

The main study was conducted between April and May 2026 and involved 12 undergraduate students, all aged 18 years or older, enrolled in Computing-related programs. The learning activity focused on three OOP concepts: inheritance, polymorphism, and abstract classes. These topics were selected because they were scheduled to be covered in the participants' course, ensuring immediate academic relevance. The experiment was hosted on the Searchat Behavior² platform [Joel et al., 2025, Machado et al., 2024], which

²Searchat Behavior: <https://github.com/lapic-ufjf/searchat-behavior>

supported task configuration, questionnaire administration, interaction logging, and configuration export for replication. Across both experimental conditions, OpenAI's GPT-4.1 served as the underlying language model.

The *Traditional* condition used the model without additional pedagogical prompting, providing direct responses without explicit instructional scaffolding. The *Socratic* condition employed the following custom system prompt:

You are a Socratic tutor. Your goal is to help the student think for themselves, guiding the conversation naturally.

General Guidelines:

- * Prioritize guiding the student to build their reasoning skills.
- * Prioritize questions that stimulate reflection and connect directly to what the student has said.
- * Avoid generic or out-of-context questions.
- * Adapt to the student's level.
- * Ensure that the key concepts of the subject are covered.

Important Priority:

- * If the student asks a direct question, answer clearly first.
- * Only then, if it makes sense, ask a question to delve deeper.

Pedagogical Diagnosis:

- * Ask the student about their prior knowledge before explaining.
- * Always assess whether the student's question depends on more basic concepts.
- * If necessary, redirect the focus to these prerequisites before explaining.
- * Use transitions such as: "'Before understanding X, we first need to understand Y'".

When to explain:

- * When the student asks.
- * When there is confusion.
- * When the conversation stalls.

How to explain:

- * Be didactic.
- * Use analogies when useful, but without forcing.

Conversation flow:

- * Prioritize naturalness over rigid rules.
- * Don't follow a fixed script; react to the context.
- * Don't ask questions if the student is still confused, clarify first.

Balance:

- * Combine questions and explanations in a fluid, not mechanical way.

Participants registered, provided informed consent, and completed a pre-questionnaire on demographics and self-reported prior knowledge. Prior knowledge was assessed using three self-assessment questions (one per target concept) rated on a Likert scale from -2 to $+2$. Most students (67%) reported limited prior knowledge. On the aggregated scale (from -6 to $+6$), the average score was -2.6 , with comparable levels between conditions (Traditional: $M = -2.7$, $SD = 2.3$; Socratic: $M = -2.5$, $SD = 2.5$),

consistent with most participants taking the course for the first time.

Participants were randomly assigned to either the *Traditional* or *Socratic* condition and instructed to use the chatbot as their primary learning resource throughout the activity. After completing the learning task, they were redirected to the immediate post-test. Participants completed a delayed retention assessment one week later. Although participants were informed that their learning would be evaluated, they were not told the content or format of the assessments in advance. The delayed retention test was not announced beforehand to better capture naturally retained knowledge and reduce expectancy effects.

The experiment was conducted in a classroom during a 2-hour session. Students were instructed to use only the provided system for learning, without consulting external resources, and to reserve at least 30 minutes for the final assessment. At the end of the session, participants completed an immediate 10-question post-test designed to assess both conceptual understanding and the ability to apply the target concepts in object-oriented software development, followed by a questionnaire regarding their experience using the chatbot as a learning tool. One week later, participants completed a 10-question knowledge retention test. Both the immediate and delayed assessments were scored using the same three-point rubric, where 0 indicated an incorrect answer, 1 a partially correct answer, and 2 a fully correct answer. Final scores were normalized to the $[0, 1]$ interval according to $Score = \frac{1}{10} \sum_{i=1}^{10} \frac{q_i}{2}$ where q_i denotes the score assigned to question i .

The study adopted a mixed-methods design, combining quantitative and qualitative analyses of learning outcomes, information behavior, and user experience. Given the limited sample size, quantitative comparisons were treated as exploratory and conducted using non-parametric methods, which do not assume normal distributions. For learning outcomes, between-group differences (Traditional vs. Socratic) were examined using the Mann-Whitney U test, within-group changes between immediate and retention assessments using the Wilcoxon signed-rank test, and effect sizes for between-group comparisons were estimated using Cliff's δ .

To assess information behavior, we analyzed participants' interaction logs considering session duration and the total number of submitted prompts. From these measures, we derived additional indicators, including prompts per minute as a proxy for interaction intensity, and immediate post-test scores per prompt and per minute as exploratory measures of interaction efficiency. Given the limited sample size, between-group comparisons for these variables were also conducted using the Mann-Whitney U test, complemented by Cliff's δ effect sizes. In addition to the quantitative analysis, the study employed a qualitative analysis to examine information-behavior patterns in participants' interactions with the chatbot. Session transcripts were systematically coded using an inductively constructed coding scheme to identify recurring interaction patterns. The resulting scheme comprises seven interpretive codes organized around two main axes: cognitive engagement and interaction regulation. Finally, user experience was analyzed through descriptive thematic analysis of four open-ended questions covering users' overall experience, the perceived adequacy of the knowledge provided for the post-questionnaire, impressions of the chatbot's interaction strategy, and additional comments.

5. Results and Discussion

This section presents the results and discussion according to the learning outcomes, information behavior, and user experience.

5.1. Learning Outcomes

Figure 1 presents the immediate and delayed retention scores for each participant, together with the group means and the average individual change ($\bar{x}\Delta$). In the immediate post-test, the *Traditional* condition achieved a higher average score than the *Socratic* condition ($M = 0.625$ vs. $M = 0.492$). After one week, however, the Traditional group's average score decreased to $M = 0.458$ ($SD = 0.166$), whereas the Socratic group showed a comparable average score ($M = 0.508$, $SD = 0.186$). This pattern is reflected in the average individual change, which was negative for the Traditional condition ($\bar{x}\Delta = -0.167$) and close to zero for the Socratic condition ($\bar{x}\Delta = +0.017$). Individual trajectories provide additional descriptive context: five of the six participants in the Traditional condition exhibited lower delayed scores, whereas the Socratic condition displayed more heterogeneous trajectories, including improvements, stable performance, and declines. Given the small sample size, these individual patterns should be interpreted only as descriptive illustrations of the observed group-level changes.

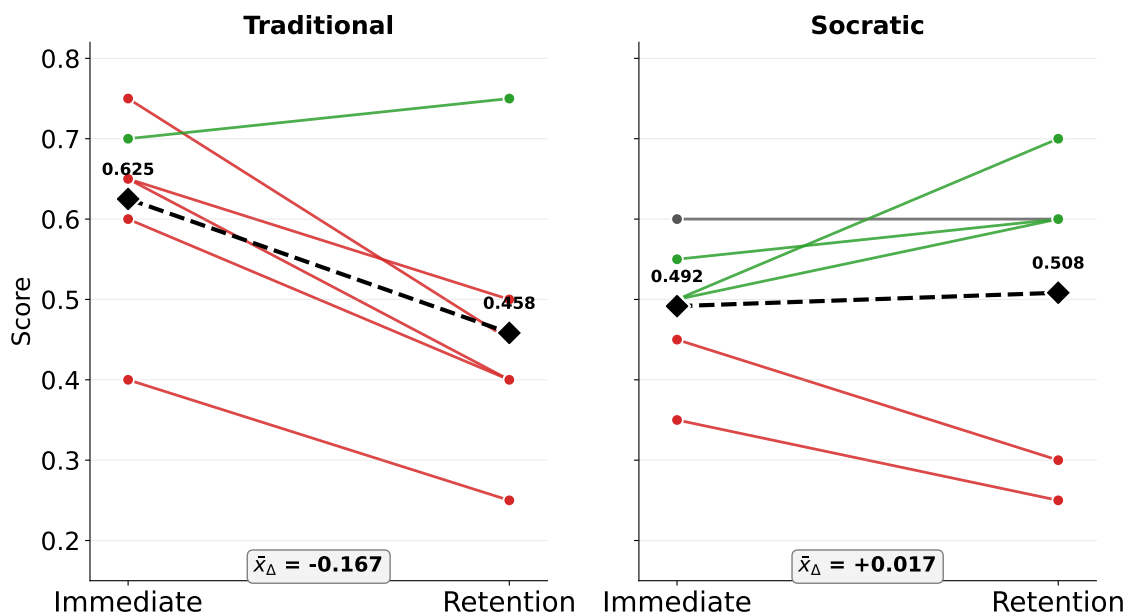


Figure 1. Individual learning trajectories from immediate to retention.

A Mann–Whitney U test comparing immediate post-test scores indicated a near-significant difference between conditions ($U = 30.5$, $p = 0.053$), favoring the *Traditional* condition, with a large effect size (Cliff's $\delta = 0.694$). In contrast, no statistically significant difference was observed in the retention test ($U = 14.5$, $p = 0.627$), with a small effect size (Cliff's $\delta = -0.194$). Within-group comparisons using the Wilcoxon signed-rank test indicated a decrease in performance for the *Traditional* group between the immediate and delayed assessments ($W = 1.0$, $p = 0.063$), whereas no meaningful change was observed for the *Socratic* group ($W = 6.5$, $p = 0.786$). Although these findings should be interpreted with caution given the limited sample size, they suggest distinct learning dynamics between the two interaction strategies: the *Traditional* condition appears to support stronger immediate performance, whereas the *Socratic* condition may contribute to more stable knowledge retention.

5.2. Information Behavior

Information behavior was analyzed using both quantitative and qualitative approaches.

5.2.1. Quantitative Information Behavior

Information behavior also differed between conditions, as summarized in Table 1. Participants in the *Traditional* condition interacted with the chatbot for an average of 36.98 minutes and submitted 13.83 prompts, whereas participants in the *Socratic* condition spent more time interacting (43.29 minutes) and submitted more prompts (18.33). To further characterize interaction intensity, we computed prompts per minute. Participants in the *Socratic* condition interacted more actively (0.42 prompts/min) than those in the *Traditional* condition (0.37 prompts/min). We estimated learning efficiency in two ways: immediate post-test performance per prompt and per minute of interaction. In both cases, the *Traditional* condition exhibited higher efficiency, suggesting that participants exposed to direct responses achieved stronger short-term performance with less interaction effort.

Table 1. Information behavior by condition.

Condition	Time (min)	# Prompts	Prompts/min	Score/Prompt	Score/Min
Traditional	36.98	13.83	0.37	0.045	0.0169
Socratic	43.29	18.33	0.42	0.027	0.0114

Mann-Whitney U tests indicated no statistically significant differences between conditions in total interaction time ($U = 12.0$, $p = 0.485$) or prompt count ($U = 10.0$, $p = 0.310$). Similarly, no statistically significant difference was observed for interaction intensity measured as prompts per minute ($U = 11.0$, $p = 0.394$). However, descriptive measures consistently suggested greater engagement in the *Socratic* condition, reflected in longer interaction sessions, more prompts, and higher interaction intensity.

Taken together, these findings suggest distinct information behavior induced by chatbot interaction strategy. The *Traditional* condition appears to have encouraged more direct information-acquisition behavior, characterized by shorter, less intensive interactions, coupled with stronger immediate performance and greater efficiency. In contrast, the *Socratic* condition appears to have promoted a more exploratory and reflective interaction pattern, characterized by longer sessions, greater conversational engagement, and more stable retention over time. It suggests that the interaction strategy may influence not only learning outcomes but also the information behavior through which learning occurs.

5.2.2. Qualitative Information Behavior

The coding scheme was developed inductively by the two first authors, who coded each transcript turn by turn, first independently and then comparing interpretations. Emerging codes were merged and refined through systematic comparison across turns and participants; ambiguous cases were resolved by consensus, and discrepancies between data and code definitions prompted revisions to the criteria rather than forced classifications. Table 2 summarizes the seven resulting interpretive codes.

Table 2. Qualitative coding scheme

Code	Definition
Active Elaboration	Learner produces new content from received information, such as rephrasing, generating examples, or connecting concepts
Difficulty Signaling	Learner indicates lack of understanding, either locating the specific point of confusion or not
Unelaborated Acknowledgment	Learner signals receipt or understanding without producing new content
Method Resistance	Learner refuses or questions the model's pedagogical strategy rather than its content
Session Regulation	Learner takes or cedes control over the direction of the conversation
Explicit Metacognition	Learner monitors and verbalizes their own learning process
Instrumental Use	Learner uses the model as a punctual technical reference rather than as a sequential tutor

The qualitative analysis revealed distinct interaction patterns across conditions. Table 3 summarizes the distribution of interpretive codes based on 74 user turns in the *Traditional* condition and 93 in the *Socratic* condition.

Table 3. Interpretive code frequencies by condition

Code	T	%T	S	%S
Active Elaboration	34	45.9	33	35.5
Session Regulation	24	32.4	42	45.2
Difficulty Signaling	15	20.3	13	14.0
Instrumental Use	14	18.9	12	12.9
Method Resistance	10	13.5	19	20.4
Explicit Metacognition	10	13.5	7	7.5
Unelaborated Acknowledgment	5	6.8	9	9.7

T = 74 turns (Traditional); S = 93 turns (Socratic)

Active Elaboration was the most frequent code in both conditions (45.9% vs. 35.5%), suggesting that active content production was not exclusive to either interaction style. The richest elaboration episodes emerged from reading code excerpts, in which learners formulated hypotheses, connected concepts, and corrected their own work.

The most pronounced contrast was observed in *Session Regulation* (32.4% vs. 45.2%). Learners in the *Socratic* condition frequently redirected the conversation, specified response formats, or ceded control to the model, reflecting active negotiation over the interaction. A similar trend was observed for *Method Resistance* (13.5% vs. 20.4%). In the *Socratic* condition, learners typically bypassed the model's questions and rephrased their requests without explicitly rejecting the strategy, whereas in the *Traditional* condition, resistance mainly took the form of conceptual critique, with learners questioning the

insufficiency or inaccuracy of the explanations received.

Conversely, *Difficulty Signaling* (20.3% vs. 14.0%) and *Explicit Metacognition* (13.5% vs. 7.5%) were more frequent in the *Traditional* condition. Two *Traditional* participants showed persistent difficulty with no observable uptake, suggesting direct content delivery was not always sufficient to promote understanding. The lower metacognition in the *Socratic* condition may reflect that the model assumed part of this regulatory function through verification questions.

Taken together, these patterns suggest that the two interaction styles were associated with qualitatively distinct behavioral profiles: the *Traditional* condition favored active elaboration and explicit metacognition at the cost of greater difficulty signaling, while the *Socratic* condition promoted learner-driven session regulation and method resistance, reflecting a more negotiated dynamic between learner and model.

5.2.3. User Experience

Overall, participants in both conditions evaluated the chatbot positively. However, thematic analysis revealed qualitative differences that were organized into three comparative categories: *perceived pedagogical role of the chatbot*, *perceived nature of the interaction* and *attribution of responsibility for learning*.

The first category, *perceived pedagogical role of the chatbot*, distinguishes how each group characterized the tool's function. *Traditional* condition participants described the chatbot predominantly in functional terms, highlighting speed, objectivity, and clarity: “*quick and straightforward*” (T1) and “*answered my questions directly with simple examples*” (T4). In contrast, *Socratic* condition participants perceived a more active pedagogical role: “*the chat responses guided me in a very good way toward learning*” (S1) and “*the interaction felt like a conversation with a teacher*” (S6).

The second category, *perceived nature of the interaction*, complements this pattern. In the *Traditional* condition, the interaction was described as transactional and similar to other available conversational systems: “*extremely common, as good as GPT, Gemini, etc.*” (T5). In the *Socratic* condition, participants perceived a more dynamic and adaptive interaction: “*at the end of each response, it always asked a question to check if I had really understood*” (S1) and “*the AI adapted to the language style*” (S6).

The third category, *attribution of responsibility for learning*, revealed the most notable contrast between conditions. In the *Traditional* condition, three participants spontaneously framed learning quality as dependent on the user's ability to formulate questions: “*it is entirely the user's responsibility [. . .] they must know the best ways to use prompts*” (T2), “*it depends more on asking the right questions*” (T5), and “*you need to know at least a little about what you are studying*” (T6). In the *Socratic* condition, this perception did not emerge; the only participant who raised the topic (S3) instead called for a more directive, system-guided structure: “*I believe it should guide me through topics pre-established by the educator*”.

Finally, contrasting these perceptions with the measured outcomes reveals a noteworthy mismatch: although participants in both conditions evaluated the chatbot positively and perceived it as an adequate learning support, overall assessment scores were modest, and the *Traditional* group in particular declined in delayed retention. This gap

between perceived and actual learning is consistent with concerns that the fluency and immediacy of AI-generated explanations may create a false sense of mastery [Gerlich, 2025, Lee et al., 2025], leading learners to overestimate their understanding [Barcaui, 2025]. It also reinforces the need to complement self-reported experience with objective learning measures when evaluating AI-based learning tools.

6. Concluding Remarks

This work investigated how different AI chatbot interaction strategies influence learning within the context of S&CAL. Motivated by concerns that direct-answer AI systems may create a false sense of mastery, this work proposed a Socratic chatbot grounded in Desirable Difficulties theory and compared it against a traditional answer-oriented chatbot in the self-directed learning of OOP concepts.

Exploratory findings suggest a trade-off between immediate performance and retention stability. Participants interacting with the Traditional chatbot achieved higher immediate performance and greater interaction efficiency, whereas those using the Socratic chatbot exhibited more stable knowledge retention over time. Qualitative analyses complement these findings by showing that the two interaction strategies fostered distinct learning behaviors: Traditional interactions tended to support more direct information acquisition, whereas Socratic interactions promoted greater learner regulation of the conversation, negotiation of the interaction process, and more exploratory engagement. Together, these findings indicate that chatbot interaction strategies shape not only what learners retain, but also how they engage with AI during learning. More broadly, our findings support the view that conversational AI should be understood not merely as an information delivery mechanism, but as an interaction environment whose pedagogical design shapes learners' cognitive and information behaviors.

This study has important limitations, most notably the small sample size, which limits statistical power and generalization. Nevertheless, the consistency between quantitative trends and qualitative observations suggests that the identified patterns are meaningful and warrant further investigation.

Future work includes conducting larger-scale longitudinal studies, evaluating different knowledge domains, exploring hybrid interaction strategies that balance direct explanations with guided questioning, and investigating how personalization and adaptive Socratic prompting may influence engagement, cognitive effort, and long-term retention. Future studies may also explore multimodal conversational environments and investigate how conversational learning behaviors evolve over extended periods of use.

Another promising direction involves developing interaction-based learning metrics that estimate learning outcomes directly from conversational interactions. In traditional SAL research, metrics such as Degree of Knowledge Gain (DKG) attempt to estimate learning outcomes from search interaction logs [Tibau et al., 2026]. Similar metrics adapted to conversational environments could help automatically estimate the probability of meaningful learning based on features such as prompt structure, interaction depth, reflection patterns, and conversational dynamics. Such metrics would be particularly valuable for real-time adaptation of AI tutors, allowing systems to dynamically adjust their pedagogical strategies according to learners' interaction patterns.

Ethical Practices

All participants were undergraduate students who voluntarily agreed to take part in the study and signed an informed consent form prior to the experiment. No sensitive personal data were collected, and participant identities were anonymized in all analyses and reported results. Furthermore, the experiment was treated as an extra activity, in which the concepts studied using AI tools were subsequently reviewed in the course's classes.

Statement on the use of Artificial Intelligence

The authors used AI-based tools to assist with language refinement and grammar correction. All scientific content, methodological decisions, interpretations, and conclusions were independently developed, reviewed, and approved by the authors, who assume full responsibility for the final content.

References

- Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open*, 12:102287.
- Bjork, E. L., Bjork, R. A., et al. (2011). Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2(59-68):56–64.
- Bjork, R. A. (1994). Memory and Metamemory Considerations in the Training of Human Beings. *Metacognition: Knowing about knowing*, 185(7.2):185–205.
- Blasco, A. and Charisi, V. (2024). AI Chatbots in K-12 Education: An Experimental Study of Socratic vs. Non-Socratic Approaches and the Role of Step-by-Step Reasoning. *Non-Socratic Approaches and the Role of Step-by-Step Reasoning (December 02, 2024)*.
- Boghossian, P. (2006). Socratic Pedagogy, Critical Thinking, and Inmate Education. *Journal of Correctional Education*, pages 42–63.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3):354–380.
- Chi, M. T. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in cognitive science*, 1(1):73–105.
- Degen, B. (2025). Resurrecting Socrates in the Age of AI: A Study Protocol for Evaluating a Socratic Tutor to Support Research Question Development in Higher Education. *arXiv preprint arXiv:2504.06294*.
- Duong, T. T. M., Da, C., and Hanh, N. (2024). The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. *Frontiers in Education*, 9:1328769.
- Fakour, H. and Imani, M. (2025). Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills. In *Frontiers in Education*, volume 10, page 1528603. Frontiers Media SA.
- Favero, L., Pérez-Ortiz, J. A., Käser, T., and Oliver, N. (2024). Enhancing Critical Thinking in Education by Means of a Socratic Chatbot. In *International workshop on AI in education and educational research*, pages 17–32. Springer.
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1):6.

- Hoppe, A., Yu, R., and Liu, J. (2026). Report on International Workshops on Investigating Learning During Web Search. *SIGIR Forum*, 59(2):1–8.
- Joel, H. d. O., Neto, A., Rosado, B., Machado, M., de Souza, J. F., and Siqueira, S. W. (2025). A Framework for Search as Learning Experiments: Design, Implementation, and Usability Insights. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 304–315. SBC.
- Kasneci, E., Sessler, K., Kühnberger, K.-U., Bannert, M., Dementieva, D., Fischer, F., and Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274.
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., and Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems*, pages 1–22.
- Lee, H.-Y., Chen, P.-H., Huang, Y.-M., and Wu, T.-T. (2026). From Immediate Answers to Structured Inquiry: A Socratic Method-Enhanced ChatGPT System for Blended Learning Discussions. *Journal of Educational Computing Research*, page 07356331261429306.
- Machado, M., Assis, E. C., Souza, J. F., and Siqueira, S. W. M. (2024). A framework to support experimentation in the context of cognitive biases in search as a learning process. In *Proceedings of the 20th Brazilian Symposium on Information Systems*, pages 1–9.
- Machado, M. d. O. C., de Alcantara Gimenez, P. J., and Siqueira, S. W. M. (2020). Raising the Dimensions and Variables for Searching as a Learning Process: A Systematic Mapping of the Literature. *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1393–1402.
- Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., and Sharma, H. (2024). A Systematic Review of Generative AI for Teaching and Learning Practice. *Education Sciences*, 14(6).
- Paul, R. (1993). *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World*. Foundation for Critical Thinking.
- Paul, R. and Elder, L. (2019). *The thinker's guide to Socratic questioning*. Bloomsbury Publishing PLC.
- Reznitskaya, A. and Gregory, M. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education*, 39(1):29–48.
- Risko, E. F. and Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9):676–688.
- Roediger, H. L. and Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3):249–255.
- Sparrow, B., Liu, J., and Wegner, D. M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science*, 333(6043):776–778.
- Tibau, M., Silva, R., Siqueira, S., and Nunes, B. (2026). Evaluating Knowledge Gain in Search Environments: An Exploratory Study of Learning Measurement. In *Anais do XXII Simpósio Brasileiro de Sistemas de Informação*, pages 478–496, Porto Alegre, RS, Brasil. SBC.
- Tibau, M., Siqueira, S. W. M., and Nunes, B. P. (2024). ChatGPT for chatting and searching: Repurposing search behavior. *Library & Information Science Research*, 46(4):101331.

- Yang, Y., Urgo, K., Arguello, J., and Capra, R. (2025). Search+Chat: Integrating Search and GenAI to Support Users with Learning-Oriented Search Tasks. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*, pages 57–70.
- Yu, R. and Liu, J. (2025). Chat as Learning in Interactive Information Retrieval and Generation. *SIGIR Forum*, 59(1):1–19.