

Learner Engagement State Typologies in AI Tutoring: A Clustering Analysis of Dialogue Behaviors in Introductory Programming Sessions

January F. Naga*

Department of Information Technology, College of Computer Studies, MSU-Iligan Institute of Technology, Philippines
E-mail: january.naga@g.msuiit.edu.ph
ORCID iD: <https://orcid.org/0000-0002-2977-6377>
*Corresponding author

Julieto E. Perez

Department of Computer Science, College of Computer Studies, MSU-Iligan Institute of Technology, Philippines
E-mail: julieto.perez@g.msuiit.edu.ph
ORCID iD: <https://orcid.org/0000-0003-3069-3883>

Received: July 5, 2025; Revised: January 17, 2026; Accepted: February 14, 2026; Published: June 8, 2026

Abstract: As Large Language Models (LLMs) become increasingly embedded in intelligent tutoring systems (ITSs), a growing need exists to understand how learners engage with these tools, especially in cognitively demanding domains such as computer programming. While prior research has focused mainly on LLM-generated scaffolding, less attention has been paid to student-side engagement, including how learners think, respond, and regulate their learning during natural language tutoring sessions. This study addresses that gap by identifying learner engagement state profiles based on behavioral and metacognitive patterns in student dialogue. Drawing on data from 36 recorded LLM-mediated C introductory programming tutorials, 1,046 dual-annotated student utterances were analyzed using K-means clustering. The analysis revealed three distinct learner engagement state profiles: Passive Reactors, Clarification Seekers, and Reflective Performers. These engagement states differed in response accuracy, metacognitive expression, and interaction style. Passive Reactors showed low initiative and limited self-regulation; Clarification Seekers demonstrated moderate accuracy and reactive help-seeking; Reflective Performers exhibited strategic engagement and high metacognitive activity. This study introduces an exploratory, scalable approach to learner profiling through natural language dialogue, advancing the design of adaptive, learner-aware LLM tutoring systems. The findings support the development of real-time learner modeling techniques that move beyond correctness, offering actionable insights for delivering more personalized and effective AI-assisted instruction in programming education.

Keywords: Large Language Models, Intelligent Tutoring Systems, Engagement State Modeling, Self-Regulated Learning, Metacognition

1. Introduction

The increasing integration of Large Language Models (LLMs) into intelligent tutoring systems (ITSs) represents a significant evolution in the design of personalized and adaptive learning environments. Unlike rule-based ITSs, LLM-powered systems support real-time, open-ended natural language interaction, enabling context-aware assistance that more closely resembles human tutoring [1, 2]. This capability has opened new opportunities for instructional support in complex domains such as computer programming, where learners often struggle with syntax precision, debugging strategies, and conceptual abstraction [3, 4].

Prior research has examined how LLMs such as GPT-3.5 and PaLM align their responses with cognitive development frameworks, including Bloom's Revised Taxonomy. For example, Lubbe, Marais, and Kruger [5] and Kasneci et al. [6] found that while LLMs effectively support lower-order cognitive processes such as remembering and understanding, they are less consistent in scaffolding higher-order thinking skills, including analysis and creation. Other studies have explored the instructional potential of LLMs to scaffold problem-solving through techniques such as prompting, clarification, and worked examples [7, 8].

However, much of this work has focused primarily on tutor-side behavior—how the LLM responds, what content it delivers, and how feedback is structured. Comparatively less attention has been paid to learner-side engagement, particularly how students interact with and respond to tutor scaffolds during real-time dialogue. As highlighted by Zawacki-Richter et al. [9] and Gašević et al. [10], understanding learner behavior in AI-mediated learning environments is essential for developing systems that adapt not only to correctness but also to learners' cognitive and metacognitive states.

Novice programmers frequently encounter cognitive and conceptual challenges when learning to code. These challenges include understanding abstract programming constructs, coordinating syntax and logic, and developing effective debugging strategies [11]. Beginners often rely on surface-level memorization rather than deeper structural reasoning [12], and these difficulties are frequently exacerbated in self-paced or online learning contexts where timely feedback may be limited [13]. Programming instruction, therefore, requires support mechanisms that address not only technical correctness but also metacognitive processes such as self-monitoring, help-seeking, and reflection [14]. LLM-powered tutoring systems offer promising tools for providing such support, yet how novice learners engage with these systems during natural language interaction remains underexplored.

Self-regulated learning (SRL) theory provides a useful framework for interpreting learner engagement in ITS environments. Prior research has shown that SRL behaviors—including help-seeking, self-monitoring, and goal-setting—are strongly associated with learning outcomes [15, 16]. In intelligent tutoring contexts, these behaviors can be inferred from behavioral and linguistic cues embedded in interaction data, such as clarification requests, reflective statements, or passive disengagement [15, 17, 18]. Azevedo et al. [19] emphasized the importance of modeling cognitive, affective, metacognitive, and motivational (CAMP) processes using multimodal data sources to support adaptive learning systems. Related work has also applied clustering techniques to interaction data to uncover recurrent patterns of engagement and strategy use [20, 21].

Despite these advances, few studies have systematically examined dialogue-based engagement patterns within LLM-powered tutoring environments using authentic student–tutor conversations. Unlike clickstream-based or quiz-driven platforms, LLM tutors operate through natural language dialogue, where behavioral and metacognitive cues are embedded in open-ended utterances rather than discrete actions. Modeling engagement in this context requires methods capable of capturing the content and intent of learner messages as they express confusion, curiosity, initiative, or reflection during tutoring interactions.

This study addresses this gap by applying clustering techniques to annotated student utterances collected from LLM-mediated introductory C programming tutoring sessions. Rather than categorizing learners as fixed types, the study aims to identify recurrent engagement state profiles—patterns of behavioral and metacognitive engagement observable at the utterance level. Specifically, this study addresses the following research questions:

- What engagement state profiles emerge from student dialogue in LLM-powered tutoring sessions?
- How do these profiles differ regarding response accuracy, metacognitive behavior, and engagement style?

By addressing these questions, this study contributes to the literature in three ways. First, it introduces an interpretable and scalable methodology for identifying engagement state profiles from natural language interactions. Second, it extends self-regulated learning theory into LLM-powered tutoring contexts by demonstrating how cognitive and metacognitive engagement states can be inferred from dialogue. Third, it provides a foundation for future research on engagement-aware tutoring systems by highlighting how interaction-level signals may inform adaptive instructional design.

This study focuses on novice learners engaged in introductory C programming and interprets findings within this context. As such, the results should not be assumed to generalize to advanced learners, non-programming domains, or alternative instructional settings without further empirical validation.

2. Literature Review

2.1. Learner Modeling and Engagement Typologies

Learner modeling has long been a central concern in intelligent tutoring system (ITS) research, to characterize how learners engage cognitively, behaviorally, and affectively during instruction. Rooted in self-regulated learning (SRL) theory, early models emphasized metacognitive processes such as monitoring, planning, and self-evaluation as key drivers of learning success [16, 22]. These theoretical foundations have informed the design of adaptive systems that respond dynamically to learners' evolving needs.

Muldner and Burleson [23] highlighted the importance of affective and behavioral modeling in ITSs, advocating for the integration of multimodal signals such as facial expressions and interaction traces to infer learner states. Similarly, Azevedo et al. [19] synthesized over a decade of research on MetaTutor, demonstrating how pedagogical agents can scaffold SRL processes through real-time adaptive feedback. Their work emphasized modeling CAMP

processes—cognitive, affective, metacognitive, and motivational—using multimodal data, while also acknowledging ongoing challenges related to interpretation, scalability, and generalization.

Complementing this line of research, Jin et al. [24] examined university students' perceptions of AI tools designed to support SRL in online environments. Their findings suggest that learners view AI systems as effective for enhancing cognitive, metacognitive, and behavioral regulation, though support for motivational regulation remains more limited. Importantly, they emphasized that psychological and pedagogical design factors—such as perceived control, learner agency, and activeness—play a critical role in the effectiveness of AI-supported SRL. This underscores the need for engagement-aware design in AI-powered tutoring systems, particularly in digital and self-paced contexts.

Recent work by Huang et al. [25] further extended this perspective by examining ambient data collection in teacher preparation environments. Their findings indicate that unobtrusive tracking of interaction patterns can support SRL and provide rich insights into engagement dynamics. Together, these studies reinforce the value of analyzing fine-grained engagement traces to construct interpretable learner models, a methodological principle adopted in the present study through behavioral and metacognitive annotation of dialogue.

2.2. Large Language Models in Education

The integration of Large Language Models (LLMs) into educational settings has expanded rapidly in recent years. Originally developed for general-purpose natural language processing tasks, LLMs such as GPT-3.5, PaLM, and LLaMA have demonstrated potential in tutoring, content generation, and learner support [6, 26]. Yan et al. [27] investigated the role of LLMs in collaborative programming by framing the AI as a third agent in human–human–AI teams. Their quasi-experimental study with middle school students found that LLM-assisted collaboration reduced cognitive load and improved computational thinking, although gains in self-efficacy were not observed. These results highlight the promise of LLMs in lowering entry barriers to complex problem-solving tasks, particularly in programming education.

Other studies have explored how LLMs can extend traditional ITS architectures. Gaeta et al. [28], for example, proposed the ITS-LLM framework, in which motivational feedback is dynamically generated by an LLM to support student engagement. Their results indicate that LLM-generated feedback can be comparable to expert-authored messages in terms of personalization, although human-authored feedback still demonstrated advantages in empathy. These findings illustrate the evolving synergy between LLMs and ITSs, particularly with respect to adaptive feedback and engagement.

Lai and Lin [29] evaluated ITS-CAL, an LLM-powered tutoring system for novice programmers. Their mixed-methods study showed that moderate use of LLM-generated feedback was associated with improved performance, whereas excessive reliance—especially among learners with lower prior knowledge—could inhibit independent problem-solving. This reinforces concerns about over-scaffolding and highlights the importance of adaptive support strategies in LLM-mediated learning environments.

Rather than focusing solely on the accuracy or cognitive level of LLM-generated outputs, recent research has emphasized the importance of examining how learners interact with LLM-based tutors during authentic learning tasks. Kasneci et al. [30] argued that while LLMs can provide immediate, adaptive explanations that may support learning and reduce entry barriers, their educational impact depends heavily on patterns of learner engagement, self-regulation, and reliance. The authors emphasized that understanding learner behaviors is critical for evaluating the pedagogical effectiveness of LLM tutors. This learner-centered perspective aligns with the present study, which analyzes student–LLM dialogue to capture engagement and metacognitive cues during tutoring interactions.

2.3. Clustering Engagement Behaviors

Clustering techniques are widely used in educational data mining (EDM) to identify latent patterns in high-dimensional learning data. K-means clustering, in particular, has been applied across diverse educational contexts due to its interpretability and computational efficiency [31]. Prior studies have used clustering to analyze clickstream logs [32], forum participation [33], and longitudinal interaction data from online courses to uncover learning strategies and engagement patterns [34].

For example, one proof-of-concept study applied k-means clustering to engagement data from a flipped spreadsheet course and found that students exhibited relatively stable interaction patterns that shifted depending on lesson content and instructional context. Such findings demonstrate how clustering can reveal meaningful engagement structures without relying on predefined learner categories.

In the context of SRL, Conati and Kardan [36] showed that engagement profiles derived from interaction logs and gaze data could inform interface adaptation in ITSs. More recently, van der Graaf et al. [37] demonstrated that real-time detection of SRL behaviors can support adaptive scaffolding in digital learning environments. Their study found that students receiving scaffolds tailored to real-time behavioral cues showed higher compliance and improved performance compared to those receiving generic support.

Building on this body of work, the present study applies clustering techniques to behavioral and metacognitive indicators extracted from student utterances in LLM-powered tutoring sessions. By focusing on utterance-level

engagement signals embedded in natural language dialogue, this study addresses a gap in the literature related to engagement-aware modeling and personalization in LLM-mediated learning environments.

3. Methodology

This study employed a data-driven methodology to identify engagement state profiles based on behavioral and metacognitive patterns embedded in student dialogue with a Large Language Model (LLM)-powered tutoring system. The analysis focused on utterances produced during introductory C programming tutorials, where students interacted with an LLM in real time to solve coding problems, ask questions, and receive feedback. The overall process followed a structured pipeline comprising five stages: data collection, annotation, feature engineering, clustering, and cluster interpretation (Figure 1).

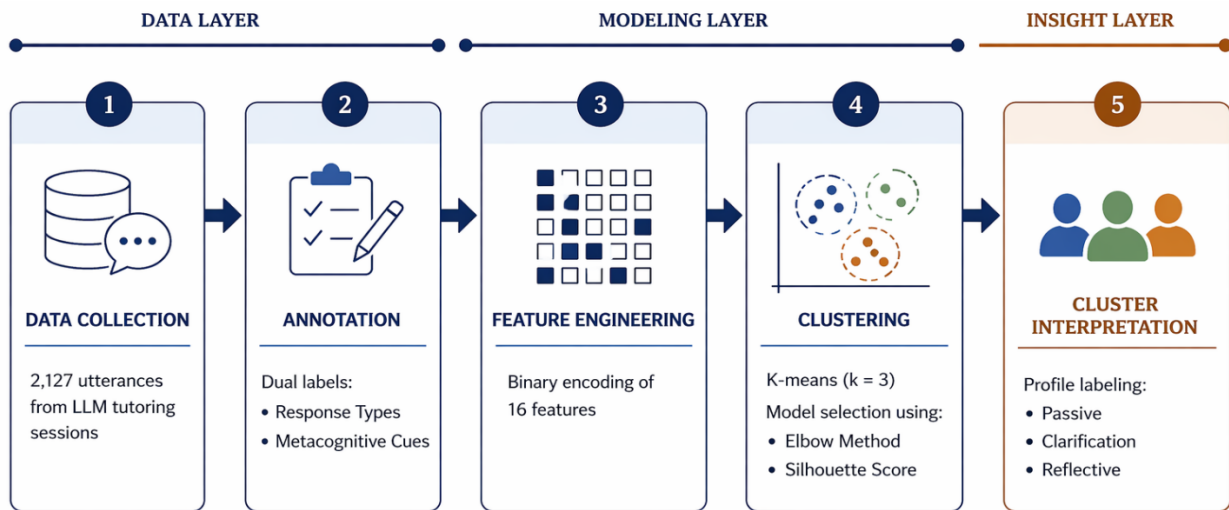


Fig. 1. Data Processing Pipeline for Engagement State Modeling

3.1. Data Collection

A total of 2,127 student utterances were collected from 36 LLM-mediated introductory C tutoring sessions. Each session involved a single student interacting with a dialogue-based tutoring environment powered by GPT-3.5, with full dialogue logging enabled under structured conditions.

Of the collected utterances, 1,046 were retained for analysis based on completeness and the presence of learning-relevant engagement behavior. Utterances were excluded if they were ambiguous, trivial, or off-topic.

Although the dataset comprised only 36 tutoring sessions, the analytic unit of this study was the utterance-level behavioral trace, yielding 1,046 annotated instances. This level of granularity aligns with prior educational data mining and intelligent tutoring system research, where learner modeling relies on fine-grained interaction events rather than participant counts [19, 36]. The resulting dataset, therefore, provides sufficient resolution for exploratory engagement state modeling despite the limited number of participants.

3.2. Annotation Process

To balance scalability and interpretive accuracy, the study employed a human-in-the-loop hybrid annotation workflow. An automated rule-based script was used to generate preliminary labels, which were subsequently reviewed and finalized by trained human coders. The automated stage functioned as a productivity aid rather than a decision authority.

Each utterance received two labels: a Response Type and a Metacognitive Cue, based on predefined linguistic heuristics (Table 1). Human coders reviewed all automated labels to ensure contextual accuracy.

Table 1. Engagement Features Used for Clustering.

Category	Features
Student Response Types	Clarification Request, Confusion, Correct Response, Error, General Question, Incorrect Response, Misconception, Off-Topic Response, Partially Correct Response
Metacognitive Cues	Clarification-Seeking, Direct Answer, Direct Response, Initiative-Taking, Passive Engagement, Request for Example, Self-Reflection

The rule-based classification script scanned each utterance for indicative keywords and syntactic patterns as preliminary signals rather than exhaustive decision rules. For example, utterances starting with “What...” or containing question marks were labeled as Clarification Requests, while those expressing uncertainty (e.g., “I think...”, “not sure...”) were flagged for Self-Reflection. Syntax-related mistakes, such as “int x = 5 +;” triggered Error annotations.

An illustration of the automated annotation logic is shown in Figure 2.

```

1 def classify_response_type(text):
2     text = text.lower()
3     if "?" in text or text.startswith("what") or text.startswith("how"):
4         return "Clarification Request"
5     elif "example" in text:
6         return "i think" or "not_sure" in text:
7     elif "wrong" in text or "error" in text:
8         return "Incorrect Response"
9     elif ";" in text and "+" in text:
10        return "Error"
11    elif "don't get" in text or "confused" in text:
12        return "Confusion"
13    else:
14        return "Direct Response"

17 def classify_metacognitive_cue(text):
17    text = text.lower()
18    if "i think" in text or "not_sure" in text:
19        return "Self-Reflection"
20    elif "can you" in text or "example" in text:
21        return "Clarification-Seeking"
22    elif "what" in text or "how" in text:
23        return "Initiative-Taking"
24    else:
25        return "Direct Response"

17 def classify_metacognitive_cue(text):
18    text = text.lower()
19    if "i think" in text or "not_sure" in text:
20        return "Self-Reflection"
21    elif "can you" in text or "example" in text:

```

Fig. 2. Rule-Based Annotation Logic for Response Types and Metacognitive Cues.

All labels were manually verified, and ambiguous cases were resolved through coder consensus. A subset of 200 utterances was double-coded, yielding a Cohen’s κ of 0.87, indicating strong inter-rater reliability.

3.3. Feature Engineering

Each annotated utterance was converted into a binary feature vector with 16 dimensions representing the presence or absence of predefined behavioral and metacognitive indicators. This utterance-level representation enabled scalable, transparent, and interpretable analysis of engagement patterns.

The feature set comprised student response types and metacognitive cues reflecting theoretically grounded indicators of cognitive engagement and self-regulated learning. For example, the utterance “*I think I need to use a loop here, but I’m not sure how to start*” was encoded with active flags for Clarification Request, Self-Reflection, and Partially Correct Response.

For example, the utterance “I think I need to use a loop here, but I’m not sure how to start” was encoded with active flags for Clarification Request, Self-Reflection, and Partially Correct Response. This encoding scheme allowed the analysis to capture meaningful metacognitive complexity without relying on opaque semantic representations or large language model embeddings, which can hinder interpretability and reproducibility in learner modeling research.

Binary encoding was adopted deliberately to prioritize interpretability and transparency. Such representations are widely used in exploratory educational data mining when the goal is pattern discovery rather than predictive optimization [31]. Each utterance was treated as an independent behavioral unit, reflecting the study’s focus on micro-level engagement signals rather than stable learner traits.

3.4. Clustering Algorithm

K-means clustering [38] was applied to the binary feature vectors to identify emergent engagement state profiles. K-means was selected due to its computational efficiency, interpretability, and established use in exploratory educational data mining.

The optimal number of clusters was determined using the Elbow Method and silhouette analysis across $k = 1-10$, with both methods converging at $k = 3$. Cluster labels were assigned strictly post hoc.

K-means++ initialization and Euclidean distance were employed to maintain consistency with centroid-based optimization [38]. Although the feature space is binary and sparse, Euclidean distance provides a transparent baseline commonly used in exploratory educational studies [21].

3.5. Cluster Validation and Interpretation

To mitigate confirmation bias, cluster interpretation was conducted only after unsupervised modeling was finalized. Feature distributions and centroid values were examined prior to assigning any semantic labels. Clusters were examined using feature-level distributions, centroid analysis, and manual review of representative utterances.

Although clustering was performed at the utterance level, the resulting clusters were conceptualized as engagement state typologies rather than fixed learner categories. These states represent recurrent patterns of cognitive and metacognitive behavior that may be exhibited multiple times by the same learner within or across sessions. Accordingly, the term profile refers to dialogic engagement configurations rather than stable learner traits. Each cluster was characterized according to patterns of behavioral and metacognitive engagement and interpreted as engagement state archetypes grounded in self-regulated learning theory [15, 16].

The three engagement state profiles identified were Passive Reactors, Clarification Seekers, and Reflective Performers.

- **Passive Reactors:** Showed confusion, disengagement, or off-topic utterances with minimal metacognitive signaling.
- **Clarification Seekers:** Actively engaged with the tutor through clarification questions, but demonstrated mixed accuracy and reflection.
- **Reflective Performers:** Demonstrated high metacognitive awareness, initiative-taking behavior, and strategic self-monitoring.

3.6. Validity and Reproducibility

To ensure annotation validity despite the use of automated pre-labeling, a rigorous coder verification process was implemented. A subset of 200 utterances was double-coded, yielding a Cohen's kappa coefficient of 0.87, indicating strong inter-rater reliability [39].

While the number of tutoring sessions was limited, the goal of this study was not statistical generalization but analytic generalization—to examine whether utterance-level behavioral signals can meaningfully support engagement state modeling in LLM-powered tutoring environments. This proof-of-concept orientation aligns with established intelligent tutoring system and learning analytics research, where learner modeling is grounded in fine-grained interaction traces rather than participant counts [19, 36].

Cluster validity was evaluated using the silhouette coefficient, which reached a value of 0.54, suggesting moderate but meaningful separation between the three identified engagement state clusters [40]. This value indicates that while the engagement states are distinguishable, boundaries between clusters are not sharply discrete and should be interpreted with appropriate caution.

To support reproducibility, all scripts used for annotation, feature vectorization, and clustering were version-controlled and implemented using open-source Python libraries, including pandas and scikit-learn.

This study establishes internal and construct validity of engagement state profiles based on behavioral and metacognitive dialogue features. However, external validity was not evaluated, as the derived engagement states were not correlated with independent learning outcomes such as programming assessment scores, task completion rates, or course retention. Accordingly, the present analysis focuses on engagement pattern discovery rather than outcome prediction.

Future work will evaluate whether similar engagement state structures emerge under alternative clustering algorithms (e.g., hierarchical, density-based, or model-based approaches), similarity metrics, and richer feature representations. Extensions of this framework will also integrate multi-valued and continuous indicators to model the intensity and temporal evolution of metacognitive behavior.

3.7. Ethical Considerations

All data used in this study were anonymized before analysis to ensure participant privacy. Informed consent was obtained from all student participants before participation. The study was conducted under institutional research guidelines and complies with Republic Act No. 10173 (Data Privacy Act of 2012, Philippines).

4. Results and Discussion

Through clustering analysis of 1,046 annotated student utterances from LLM-powered introductory C programming tutoring sessions, three distinct engagement state profiles emerged: Passive Reactors, Clarification

Seekers, and Reflective Performers. Each profile represents a recurrent pattern of behavioral and metacognitive engagement observable at the utterance level and was validated through feature-level distributions, cluster visualizations, and theoretical alignment with self-regulated learning frameworks [15, 16]. These engagement state profiles reflect dynamic engagement tendencies rather than fixed learner identities, allowing for intra-learner variability over time.

4.1. Engagement State Profiles Emerging from LLM-Powered Tutoring Sessions

Three engagement state clusters were identified using K-means clustering on 16 binary features, capturing both response types and metacognitive cues. In this study, learner engagement state profiles refer to recurrent patterns of learner engagement observable at the utterance level, rather than stable learner attributes or classifications. The optimal number of clusters ($k = 3$) was determined through the Elbow Method and further validated by a silhouette coefficient of 0.54, indicating moderate but interpretable separation among engagement state clusters. Importantly, engagement state labels were assigned only after statistical clustering and were not used to constrain or guide the unsupervised modeling process.

Table 2 presents the mean occurrence of each behavioral and metacognitive feature per engagement state cluster. These feature-level differences formed the basis for interpreting and characterizing each engagement state profile.

Table 2. Feature Distribution by Cluster.

Feature	Passive Reactors	Clarification Seekers	Reflective Performers
Clarification Request	0.013	0.006	0.258
Confusion	0.195	0.023	0.077
Correct Response	0.094	0.146	0.033
Error	0.015	0.010	0.015
General Question	0.021	0.003	0.273
Incorrect Response	0.092	0.146	0.041
Misconception	0.047	0.062	0.015
Off-Topic Response	0.079	0.023	0.015
Partially Correct Response	0.341	0.578	0.144
Request for Example	0.101	0.003	0.129
Clarification-Seeking	0.000	0.000	1.000
Direct Answer	0.530	0.000	0.000
Direct Response	0.000	1.000	0.000
Initiative-Taking	0.170	0.000	0.000
Passive Engagement	0.167	0.000	0.000
Self-Reflection	0.133	0.000	0.000

Cluster 1: Passive Reactors

Utterances associated with the Passive Reactor engagement state exhibited low levels of cognitive engagement and limited metacognitive regulation. These utterances frequently included off-topic responses, expressions of confusion, and partially correct answers. This engagement state was characterized by relatively high rates of passive engagement (0.167), frequent confusion (0.195), and a tendency toward direct but unelaborated responses (0.530).

The observed behavioral pattern reflects surface-level learning processes and minimal self-regulatory activity. Within the framework of self-regulated learning theory, such patterns align with low levels of self-monitoring and goal-setting, which are commonly associated with underdeveloped regulatory capacity [16]. This engagement state reflects interaction sequences marked by limited initiative and reliance on external prompting.

Cluster 2: Clarification Seekers

Utterances associated with the Clarification Seeker engagement state demonstrated moderate cognitive engagement characterized by frequent clarification requests and reactive participation. These utterances included a high proportion of partially correct responses (0.578), alongside notable occurrences of incorrect (0.146) and misconception-based responses, suggesting active involvement accompanied by incomplete conceptual understanding.

A defining characteristic of this engagement state was consistent Direct Response behavior (1.000), indicating reliance on tutor-initiated prompts rather than autonomous dialogue initiation. This interaction pattern aligns with prior descriptions of help-seeking but dependent engagement styles, in which learners actively seek guidance while remaining reliant on external scaffolds to progress through instructional tasks [41]. Collectively, the Clarification Seeker profile reflects engagement sequences that are cognitively present yet metacognitively transitional.

Cluster 3: Reflective Performers

Utterances associated with the Reflective Performer engagement state exhibited the highest levels of metacognitive awareness and strategic engagement. These utterances frequently contained clarification requests, general conceptual questions, requests for examples, and explicit self-reflective cues, indicating deep cognitive involvement and proactive regulation of learning processes.

This engagement state was characterized by high values for Clarification Request (0.258), General Question (0.273), and Request for Example (0.129). Notably, Reflective Performers were the only engagement state consistently exhibiting Clarification-Seeking cues (1.000), highlighting strong self-monitoring and intentional regulation. This profile closely aligns with the interactive mode in Chi’s Active–Constructive–Interactive framework, which is associated with deeper learning and knowledge construction [42].

Utterances within this engagement state reflect strategic use of the tutor as a cognitive resource for hypothesis testing, ambiguity resolution, and conceptual extension, consistent with advanced self-regulatory behavior.

Table 3 compares the three engagement state profiles across key dimensions of response accuracy, metacognitive behavior, and engagement style. Reflective Performers demonstrate consistently high levels across all dimensions, whereas Passive Reactors exhibit uniformly low engagement indicators. Clarification Seekers occupy an intermediate position, reflecting emerging regulatory behaviors.

Table 3. Engagement State Profile Comparison.

Dimension	Passive Reactors	Clarification Seekers	Reflective Performers
Response Accuracy	Low, off-topic, and confused responses	Moderate; many partially correct or incorrect responses	High; few incorrect responses, active questioning
Metacognitive Behavior	Minimal to none; passive engagement	Moderate, direct, and clarification-seeking responses	High self-reflection, clarification-seeking, and strategic inquiry
Engagement Style	Reactive only when prompted	Reactive, content-driven engagement	Initiative-taking and exploratory dialogue

The radar visualization shown in Figure 3 further illustrates differences across engagement state profiles, highlighting the heterogeneity of learner engagement patterns in LLM-powered tutoring environments. Although the engagement states exhibit coherent behavioral patterns, the moderate silhouette value indicates that transitions between states are likely fluid rather than sharply bounded. Accordingly, these clusters should be interpreted as probabilistic engagement tendencies rather than rigid categories, reinforcing the view of learner engagement as dynamic and context-sensitive. In line with prior intelligent tutoring systems research, real-time recognition of metacognitive signals has been identified as a critical foundation for personalizing feedback and scaffolding [19, 37].

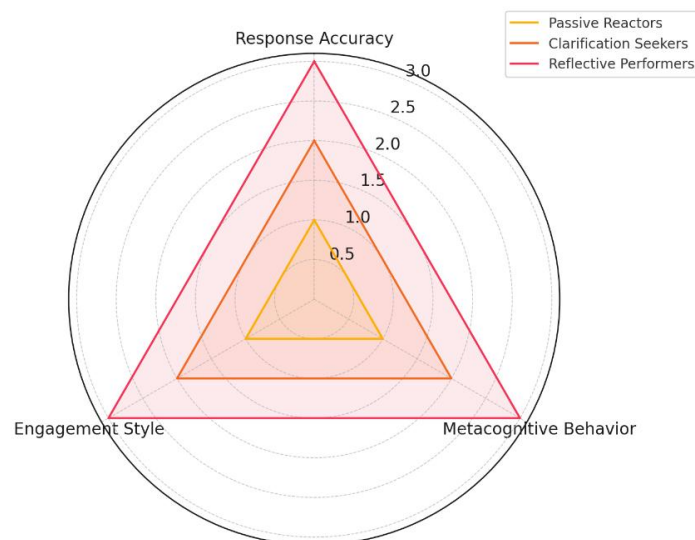


Fig. 3. Engagement State Profile Comparison.

4.2. Differences in Accuracy, Metacognitive Behavior, and Engagement Style across Learner Profiles

The three learner engagement state profiles identified in this study—Passive Reactors, Clarification Seekers, and Reflective Performers—exhibited distinct patterns in interaction frequency, response accuracy, metacognitive signaling, and engagement style. Differences across three core dimensions—response accuracy, metacognitive behavior, and engagement style—are discussed below:

1. Response Accuracy

Response accuracy varied across engagement state profiles. Utterances associated with the Clarification Seeker engagement state showed the highest proportion of partially correct responses (0.578), reflecting active engagement with the task alongside persistent conceptual uncertainty. The concurrent presence of incorrect and misconception-based responses suggests ongoing hypothesis testing and iterative refinement of understanding.

By contrast, utterances associated with the Reflective Performer engagement state exhibited fewer incorrect responses and a higher prevalence of clarification requests and general conceptual questions. This pattern indicates deliberate verification and conceptual exploration before committing to solutions, consistent with metacognitive regulation strategies that prioritize understanding over response speed.

Utterances associated with the Passive Reactor engagement state demonstrated the lowest overall accuracy. These utterances were frequently off-topic, vague, or confusion-driven and included a relatively high proportion of direct but unelaborated responses. Such patterns reflect limited cognitive investment and align with surface-level engagement processes observed in early or disengaged learning states.

These trends are consistent with prior self-regulated learning research, which associates deeper processing with delayed response strategies and superficial engagement with impulsive or minimal cognitive investment [15, 16].

2. Metacognitive Behavior

A clear gradient of metacognitive sophistication emerged across engagement state profiles. Utterances associated with the Reflective Performer engagement state exhibited robust metacognitive activity, including self-reflection, clarification-seeking, and requests for examples. These behaviors reflect intentional monitoring, evaluation, and regulation of understanding, aligning with frameworks that emphasize the integration of cognitive and metacognitive processes in effective learning [19].

Utterances associated with the Clarification Seeker engagement state also demonstrated metacognitive activity, though cues were predominantly reactive. Clarification requests and direct responses frequently followed tutor prompts or moments of confusion, suggesting engagement that is externally triggered rather than internally regulated.

In contrast, utterances associated with the Passive Reactor engagement state exhibited minimal metacognitive signaling. Clarification requests, self-reflection, and initiative-taking behaviors were rare, with interaction patterns dominated by passive engagement and off-topic remarks. This absence of self-monitoring and strategy use reflects limited metacognitive awareness within this engagement state.

3. Engagement Style

Engagement style further differentiated the profiles. Utterances associated with the Reflective Performer engagement state demonstrated initiative-taking and dialogic interaction, characterized by proactive questioning, reflective statements, and strategic use of the tutor as a cognitive resource. This engagement style aligns with Chi's interactive mode of learning, which is associated with deeper knowledge construction and conceptual integration [42].

Utterances associated with the Clarification Seeker engagement state reflected reactive, content-driven engagement. While cognitively active, interactions typically followed tutor prompts rather than initiating new inquiry, indicating reliance on external structure during learning.

Utterances associated with the Passive Reactor engagement state exhibited minimal interactional initiative. Dialogue contributions were sparse, often vague, and frequently disconnected from the task, reflecting constrained engagement and limited dialogic participation.

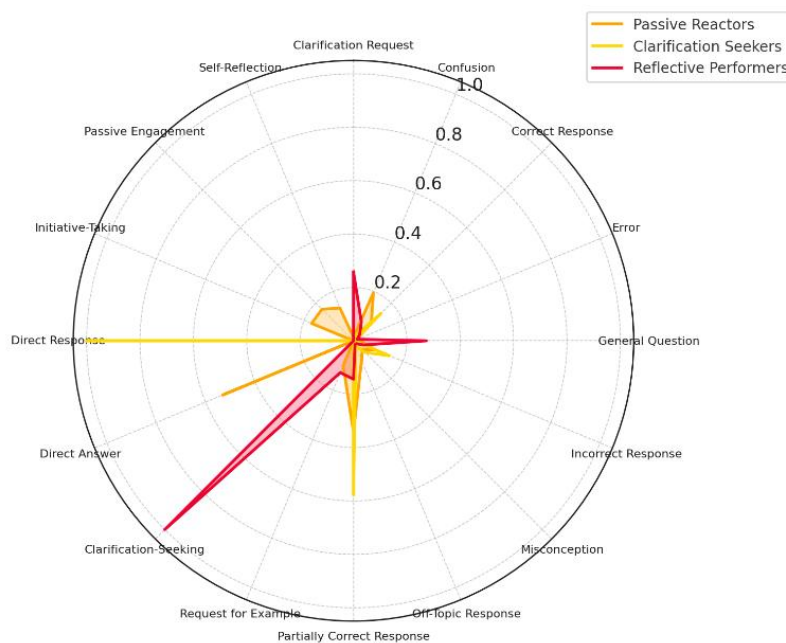


Fig. 4. Normalized Radar Chart of Engagement Intensity by Cluster Label.

To illustrate these differences, Figure 4 presents a normalized radar chart showing engagement feature intensities across all 16 behavioral and metacognitive indicators. Reflective Performers exhibit pronounced peaks in clarification-seeking, self-reflection, and requests for examples, while Clarification Seekers emphasize direct responses and partially correct answers. Passive Reactors display elevated levels of confusion, passive engagement, and off-topic behavior.

Complementing this, Figure 5 displays a grouped bar chart of raw feature frequencies, enabling direct magnitude comparisons across engagement state profiles. Together, these visualizations demonstrate that engagement in LLM-powered tutoring environments is heterogeneous and varies substantially across accuracy, metacognitive signaling, and interaction style.

Overall, these findings demonstrate that utterance-level engagement patterns provide a meaningful lens for analyzing how learners interact with LLM-powered tutors. Modeling behavioral and metacognitive signals at this granularity enables nuanced differentiation of engagement dynamics, consistent with prior intelligent tutoring systems research that has leveraged interaction-level data to uncover latent engagement patterns [21, 38].

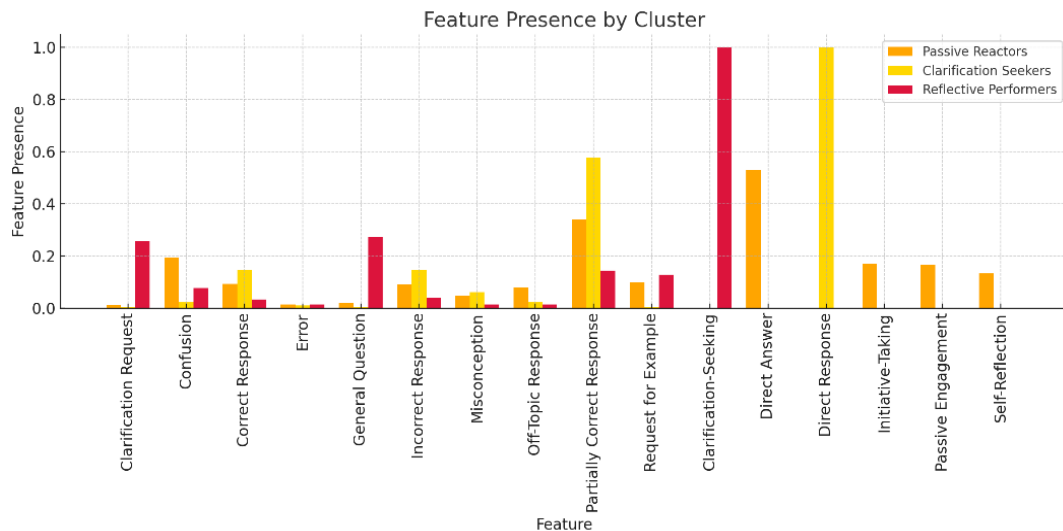


Fig. 5. Radar Chart of Engagement Feature Intensities by Cluster.

4.3. Theoretical and Practical Implications

The identification of three dialogue-based engagement states among novice C programming learners contributes initial evidence toward understanding how learner behavior manifests in LLM-powered programming tutors. While much prior work has emphasized the effectiveness of tutor-generated scaffolds aligned with frameworks such as Bloom’s Revised Taxonomy [5], comparatively few studies have systematically analyzed learner-side behaviors in response to such scaffolds, particularly within natural language interactions. This study addresses that gap by demonstrating how engagement patterns—especially metacognitive signals embedded in dialogue—serve as reliable indicators for characterizing actionable engagement state profiles. The present study is exploratory and observational in nature; accordingly, all statements regarding how specific engagement states may respond to instructional strategies are hypothesis-generating rather than causal claims. The present study is exploratory and observational in nature. Accordingly, all statements regarding how specific engagement states may respond to instructional strategies are hypothesis-generating rather than causal claims.

These findings also extend the application of self-regulated learning (SRL) theory [15, 16] into LLM-powered tutoring environments. Unlike traditional approaches that rely on think-aloud protocols or retrospective surveys, this study shows that behavioral cues embedded in student dialogue can function as real-time proxies for cognitive and metacognitive states. This approach enables the construction of engagement-aware learner models that are data-driven, interpretable, and responsive to moment-to-moment interaction dynamics.

From a practical perspective, the three engagement states differ systematically in their behavioral and metacognitive characteristics. Passive Reactors are characterized by low self-regulation and limited initiative, Clarification Seekers exhibit reactive but cognitively engaged behavior, and Reflective Performers demonstrate sustained metacognitive awareness and strategic interaction. At present, these engagement states should be interpreted as behavioral descriptors rather than predictors of learning outcomes or performance.

4.4. Design Propositions for Adaptive LLM Tutors

The following design propositions are derived from observed engagement patterns and grounded in self-regulated learning theory. They are interpreted as hypothesis-generating guidelines for future system development and evaluation rather than empirically validated instructional effects.

- *Passive Reactor engagement:* Utterances associated with the Passive Reactor engagement state may signal low cognitive initiative and disengagement. Such states may respond favorably to motivational interventions, confidence-building strategies, and structured, step-by-step guidance. Interface features such as visual cues or affective nudges could be explored as mechanisms for reducing cognitive entry barriers.
- *Clarification Seeker engagement:* Utterances associated with the Clarification Seeker engagement state indicate active but reactive participation. These states may be suitable candidates for adaptive scaffolding strategies that evolve over time, including gradual fading of direct support or increased use of Socratic questioning to encourage greater cognitive autonomy.
- *Reflective Performer engagement:* Utterances associated with the Reflective Performer engagement state exhibit high metacognitive awareness and initiative. These states may be appropriate for advanced instructional strategies, such as synthesis-based tasks, comparative reasoning prompts, or reflective activities that encourage articulation of reasoning and exploration of alternative solutions.

These propositions have implications not only for automated tutor logic but also for instructor-facing dashboards. Real-time identification of engagement states from learner dialogue may provide educators with on-the-fly diagnostics regarding student readiness for different forms of feedback or challenge. Future research should implement controlled experiments comparing adaptive and non-adaptive LLM tutoring conditions to empirically evaluate whether engagement state-aware scaffolding leads to measurable gains in learning outcomes, retention, or metacognitive growth.

Integrating engagement state-aware modeling into LLM-powered tutoring systems offers a promising direction for developing more responsive, personalized, and pedagogically aligned AI instruction, subject to future experimental validation.

4.5. Limitations and Future Work

While this study offers valuable insights into learner profiling within LLM-powered tutoring systems, several limitations should be acknowledged.

First, the relatively small number of tutoring sessions ($n = 36$) constrains the statistical generalizability of the identified engagement state profiles. Consequently, the reported clusters should be interpreted as emergent behavioral archetypes rather than population-level typologies. However, the large volume of utterance-level data (1,046 instances) and high inter-rater reliability support the internal validity of the findings. Future work should replicate the proposed framework using larger and more diverse datasets to evaluate the stability and scalability of engagement state profiles across contexts.

Because clustering was conducted at the utterance level, the identified profiles should not be interpreted as fixed learner identities. Instead, they reflect dynamic engagement states that learners may transition between during tutoring interactions. Future studies could aggregate engagement state trajectories at the learner level to derive higher-order learner typologies while preserving temporal dynamics.

Second, the dataset was restricted to utterances from a single instructional domain—introductory C programming. Although this context is pedagogically relevant, it may limit the generalizability of the findings to other domains such as mathematics, science, language learning, or to learners with different levels of expertise. Accordingly, the present engagement states should be viewed as domain- and population-specific until validated across instructional contexts.

Third, although the study employed a hybrid annotation workflow with high inter-rater reliability, the feature labeling process involved elements of human judgment. While this enhances interpretability and transparency, it introduces subjectivity. In addition, the initial reliance on keyword heuristics may introduce subtle biases, particularly for utterances with implicit intent or culturally nuanced phrasing. Future research should evaluate supervised or LLM-assisted annotation models trained on the validated dataset to improve sensitivity to pragmatic and contextual cues while maintaining interpretability.

Fourth, the present analysis treated utterances as independent observations and did not explicitly model temporal dependencies or within-learner correlations. While this design enabled detection of fine-grained engagement patterns, it does not capture how learners transition between engagement states over time or how interaction sequences evolve within a session. Future work should incorporate temporal modeling techniques (such as Hidden Markov Models, sequence mining, or mixed-effects clustering) to trace longitudinal engagement trajectories and account for within-learner dependency.

Fifth, this study established internal and construct validity of engagement state profiles based on behavioral and metacognitive dialogue features but did not evaluate external validity. The engagement states were not linked to independent learning outcomes such as assessment scores, task completion rates, or course retention. Accordingly, the present analysis focuses on engagement pattern discovery rather than outcome prediction. In addition, the use of binary

features necessarily reduces the richness of linguistic expression by collapsing nuanced utterances into categorical indicators. Future work should explore richer representations, including weighted features, ordinal scales, sentiment intensity measures, or contextual embeddings, to capture finer-grained engagement dynamics.

Sixth, the feature space was limited to binary indicators and analyzed using the K-means clustering algorithm. While this approach supported interpretability and methodological transparency, alternative modeling techniques—such as Gaussian Mixture Models, hierarchical clustering, or deep representation learning—may reveal subtler or overlapping engagement states. The reliance on Euclidean distance for sparse binary features may also attenuate sensitivity to rare but pedagogically meaningful behaviors. Future studies should compare alternative distance metrics (e.g., Jaccard, cosine similarity, Hamming distance) to assess their impact on engagement state formation and cluster stability. Although $k = 3$ was selected through quantitative validation, the semantic labeling of clusters involves interpretive judgment, and alternative cluster solutions may be plausible.

Finally, this study did not experimentally evaluate the effects of adapting LLM tutor responses based on engagement state classification. Future work should investigate real-time application of engagement-aware modeling to dynamically adjust tutor scaffolding, feedback strategies, or instructional pathways. Controlled intervention studies assessing the impact of such adaptations on learning outcomes, retention, and metacognitive growth are necessary to advance the development of intelligent, learner-aware educational systems.

5. Conclusion

This study examined how dialogue-based engagement states emerge in LLM-powered tutoring environments through clustering analysis of annotated student utterances. Using behavioral and metacognitive indicators, the analysis identified three distinct engagement state profiles—Passive Reactors, Clarification Seekers, and Reflective Performers—representing recurrent patterns of learner engagement observable at the utterance level. These engagement states differed systematically in response accuracy, metacognitive signaling, and interaction style, offering a multidimensional view of how learners engage with AI tutors during real-time instructional dialogue.

The findings highlight the value of utterance-level analysis for characterizing engagement state typologies in natural language tutoring environments. Dialogue-based features were shown to provide insight into cognitive and regulatory processes as they unfold in interaction, without relying on retrospective measures or self-report instruments. Reflective Performer engagement states exhibited the strongest alignment between metacognitive activity and response accuracy, whereas Passive Reactor states were characterized by limited initiative and self-regulatory signaling. Importantly, these engagement states should be interpreted as empirically grounded but evolving patterns of interaction rather than fixed learner categories.

This work contributes to the development of engagement-aware AI tutoring systems by presenting an interpretable, data-driven framework for modeling learner engagement in LLM-mediated education. It also bridges methodological gaps between self-regulated learning theory and contemporary AI tutoring research by demonstrating how cognitive and metacognitive engagement states can be inferred from natural language dialogue. While the present study focuses on introductory C programming with novice learners, the proposed framework provides a foundation for future investigations across domains, learner populations, and instructional formats.

Future research should prioritize replication with larger and more diverse datasets, incorporation of temporal modeling to capture engagement state transitions, and validation against independent learning outcomes. The current framework uses binary engagement features to support interpretability. It provides a baseline architecture that can be extended with richer representations such as weighted features, ordinal measures, or contextual embeddings. Linking engagement state profiles to learning outcomes through controlled intervention studies represents a critical next step toward establishing the predictive and instructional relevance of engagement-aware modeling in LLM-powered tutoring systems.

All the Declarations and Statements

Author Contributions Statement

January F. Naga – Conceptualization, Methodology, Formal Analysis, Visualization, and Writing – Original Draft: Developed the research concept, designed the methodology, conducted data analysis and interpretation, prepared visualizations, and drafted the manuscript.

Julieto E. Perez – Conceptualization, Supervision, Validation, Writing – Review and Editing, and Project Administration: Developed the research concept, provided research supervision, validated methodological and analytical procedures, reviewed and revised the manuscript for intellectual content, and supported project coordination.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding Declaration

This research received no external funding.

Data Availability Statement

The data used in this study consist of anonymized student–LLM interaction transcripts collected under controlled research conditions. Due to privacy and ethical considerations, the dataset is not publicly available.

However, processed data and analysis scripts may be made available from the corresponding author upon reasonable request.

Ethical Declarations

This study involves human participants and was conducted in accordance with institutional and national research guidelines. Informed consent was obtained from all participants prior to data collection. All data were anonymized to ensure participant confidentiality and privacy. The study complies with the provisions of the Data Privacy Act of 2012 (Republic Act No. 10173, Philippines).

Acknowledgments

The authors would like to express their sincere gratitude to the participants of this study for their valuable contributions. The authors also thank colleagues and reviewers for their constructive feedback, which significantly improved the quality of this manuscript.

Declaration of Generative AI in Scholarly Writing

No AI tools were used to generate data, perform analysis, or derive conclusions. AI tools are not listed as authors. The authors take full responsibility for the content of this manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

AI – Artificial Intelligence

NLP – Natural Language Processing

LLM – Large Language Model

ITS – Intelligent Tutoring System

SRL – Self-Regulated Learning

DL – Deep Learning

Appendix

None.

References

- [1] S. Steinert, K. E. Avila, S. Ruzika, J. Kuhn, and S. Kuchemann, "Harnessing large language models to develop research-based learning assistants for formative feedback," *Smart Learning Environments*, vol. 11, no. 1, p. 62, 2024/12/19 2024, doi: <https://doi.org/10.1186/s40561-024-00354-1>.
- [2] W. Xing, N. Nixon, S. Crossley, P. Denny, A. Lan, J. Stamper, and Z. Yu, "The Use of Large Language Models in Education," *International Journal of Artificial Intelligence in Education*, vol. 35, no. 2, pp. 439-443, 2025/06/01 2025, doi: <https://doi.org/10.1007/s40593-025-00457-x>.
- [3] G. Jošt, V. Taneski, and S. Karakatič, "The Impact of Large Language Models on Programming Education and Student Learning Outcomes," *Applied Sciences*, vol. 14, no. 10, p. 4115, 2024, doi: <https://doi.org/10.3390/app14104115>.
- [4] M. Giannakos et al., "The promise and challenges of generative AI in education," *Behaviour & Information Technology*, pp. 1-27, doi: <https://doi.org/10.1080/0144929X.2024.2394886>.
- [5] A. Lubbe, E. Marais, and D. Kruger, "Cultivating independent thinkers: The triad of artificial intelligence, Bloom's taxonomy and critical thinking in assessment pedagogy," *Education and Information Technologies*, 2025/03/12 2025, doi: <https://doi.org/10.1007/s10639-025-13476-x>.
- [6] E. Kasneci et al., "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences*, vol. 103, p. 102274, 01/30 2023, doi: <https://doi.org/10.1016/j.lindif.2023.102274>.
- [7] Z. Pardos and S. Bhandari, "ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills," *PLOS ONE*, vol. 19, 05/24 2024, doi: <https://doi.org/10.1371/journal.pone.0304013>.

- [8] P. Cui and M. Sachan, Investigating the Zone of Proximal Development of Language Models for In-Context Learning. 2025, pp. 6470-6483.
- [9] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, 2019/10/28 2019, doi: 10.1186/s41239-019-0171-0.
- [10] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *The Internet and Higher Education*, vol. 28, pp. 68-84, 2016/01/01/ 2016, doi: <https://doi.org/10.1016/j.iheduc.2015.10.002>.
- [11] E. Lahtinen, K. Ala-Mutka, and H.-M. Järvinen, A study of the difficulties of novice programmers. 2005, pp. 14-18.
- [12] R. Lister et al., "A Multi-National Study of Reading and Tracing Skills in Novice Programmers," *ACM SIGCSE Bulletin*, vol. 36, pp. 119-150, 01/01 2005, doi: 10.1145/1044550.1041673.
- [13] A. Robins, R. Janet, and N. and Rountree, "Learning and Teaching Programming: A Review and Discussion," *Computer Science Education*, vol. 13, no. 2, pp. 137-172, 2003/06/01 2003, doi: <https://doi.org/10.1076/csed.13.2.137.14200>.
- [14] P. Kinnunen and L. Malmi, "Why students drop out CS1 course?," presented at the Proceedings of the second international workshop on Computing education research, Canterbury, United Kingdom, 2006. [Online]. Available: <https://doi.org/10.1145/1151588.1151604>.
- [15] E. Panadero, "A Review of Self-regulated Learning: Six Models and Four Directions for Research," *Frontiers in Psychology*, vol. 8, 2017, doi: <https://doi.org/10.3389/fpsyg.2017.00422>.
- [16] B. J. Zimmerman, "Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner Eds. San Diego: Academic Press, 2000, pp. 13-39.
- [17] C. A. Brenner, "Self-regulated learning, self-determination theory and teacher candidates' development of competency-based teaching practices," *Smart Learning Environments*, vol. 9, no. 1, p. 3, 2022/01/06 2022, doi: <https://doi.org/10.1186/s40561-021-00184-5>.
- [18] F. Arianto and M. Hanif, "Evaluating metacognitive strategies and self-regulated learning to predict primary school students self-efficacy and problem-solving skills in science learning," *Journal of Pedagogical Research*, 2024, doi: <https://doi.org/10.33902/jpr.202428575>.
- [19] R. Azevedo et al., "Lessons Learned and Future Directions of MetaTutor: Leveraging Multichannel Data to Scaffold Self-Regulated Learning With an Intelligent Tutoring System," (in English), *Frontiers in Psychology*, Review vol. Volume 13 - 2022, 2022-June-14 2022, doi: 10.3389/fpsyg.2022.813632.
- [20] D. Sachan and K. Saroha, "A Review of Adaptive and Intelligent Online Learning Systems," Singapore, 2022: Springer Nature Singapore, in *ICT Analysis and Applications*, pp. 251-262, doi: https://doi.org/10.1007/978-981-16-5655-2_24.
- [21] Y. Fang, A. Lippert, Z. Cai, S. Chen, J. C. Frijters, D. Greenberg, and A. C. Graesser, "Patterns of Adults with Low Literacy Skills Interacting with an Intelligent Tutoring System," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 2, pp. 297-322, 2022/06/01 2022, doi: <https://doi.org/10.1007/s40593-021-00266-y>.
- [22] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in *Metacognition in educational theory and practice.*, (The educational psychology series. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 1998, pp. 277-304.
- [23] K. Muldner and W. Bursleson, "Utilizing sensor data to model students' creativity in a digital environment," *Computers in Human Behavior*, vol. 42, pp. 127-137, 2015/01/01/ 2015, doi: <https://doi.org/10.1016/j.chb.2013.10.060>.
- [24] S.-H. Jin, K. Im, M. Yoo, I. Roll, and K. Seo, "Supporting students' self-regulated learning in online learning using artificial intelligence applications," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 37, 2023/06/26 2023, doi: 10.1186/s41239-023-00406-5.
- [25] L. Huang, L. Dias, E. Nelson, L. Liang, S. P. Lajoie, and E. G. Poitras, "The Role of Self-Improving Tutoring Systems in Fostering Pre-Service Teacher Self-Regulated Learning," (in English), *Frontiers in Artificial Intelligence*, Original Research vol. Volume 4 - 2021, 2022-January-03 2022, doi: 10.3389/frai.2021.769455.
- [26] A. J. Carroll and J. Borycz, "Integrating large language models and generative artificial intelligence tools into information literacy instruction," *The Journal of Academic Librarianship*, vol. 50, no. 4, p. 102899, 2024/07/01/ 2024, doi: <https://doi.org/10.1016/j.acalib.2024.102899>.
- [27] Y.-M. Yan, C.-Q. Chen, Y.-B. Hu, and X.-D. Ye, "LLM-based collaborative programming: impact on students' computational thinking and self-efficacy," *Humanities and Social Sciences Communications*, vol. 12, no. 1, p. 149, 2025/02/07 2025, doi: 10.1057/s41599-025-04471-1.
- [28] A. Gaeta, F. Orciuoli, A. Pascuzzo, and A. Peduto, "Enhancing traditional ITS architectures with large language models for generating motivational feedback," *Computers and Education: Artificial Intelligence*, vol. 9, p. 100433, 2025/12/01/ 2025, doi: <https://doi.org/10.1016/j.caeai.2025.100433>.
- [29] C.-H. Lai and C.-Y. Lin, "Analysis of Learning Behaviors and Outcomes for Students with Different Knowledge Levels: A Case Study of Intelligent Tutoring System for Coding and Learning (ITS-CAL)," *Applied Sciences*, vol. 15, no. 4, p. 1922, 2025. [Online]. Available: <https://www.mdpi.com/2076-3417/15/4/1922>.
- [30] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, and G. Kasneci, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, Art. no. 102274, 03/01 2023, doi: 10.1016/j.lindif.2023.102274.
- [31] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. Morgan Kaufmann, 2022.
- [32] V. Kovanovic, S. Joksimovic, D. Gasevic, G. Siemens, and M. Hatala, "What public media reveals about MOOCs: A systematic analysis of news reports," *British Journal of Educational Technology*, vol. 46, pp. 510-527, 04/01 2015, doi: 10.1111/bjet.12277.
- [33] J. A. Larusson and R. Alterman, "Wikis to support the "collaborative" part of collaborative learning," *International Journal of Computer-Supported Collaborative Learning*, vol. 4, no. 4, pp. 371-402, 2009/12/01 2009, doi: 10.1007/s11412-009-9076-6.

- [34] R. Davies, G. Allen, C. Albrecht, N. Bakir, and N. Ball, "Using Educational Data Mining to Identify and Analyze Student Learning Strategies in an Online Flipped Classroom," *Education Sciences*, vol. 11, no. 11, p. 668, 2021. [Online]. Available: <https://www.mdpi.com/2227-7102/11/11/668>.
- [35] A. L'áourneau, M. Deslandes Martineau, P. Charland, J. A. Karran, J. Boasen, and P. M. L'éger, "A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education," *npj Science of Learning*, vol. 10, no. 1, p. 29, 2025/05/14 2025, doi: 10.1038/s41539-025-00320-7.
- [36] C. Conati and S. Kardan, "Student Modeling: Supporting Personalized Instruction, from Problem Solving to Exploratory Open-Ended Activities," *AI Magazine*, vol. 34, pp. 13-26, 09/01 2013, doi: 10.1609/aimag.v34i3.2483.
- [37] J. van der Graaf et al., "How to design and evaluate personalized scaffolds for self-regulated learning," *Metacognition and Learning*, vol. 18, no. 3, pp. 783-810, 2023/12/01 2023, doi: 10.1007/s11409-023-09361-y.
- [38] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178-210, 2023/04/01/ 2023, doi: <https://doi.org/10.1016/j.ins.2022.11.139>.
- [39] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159-174, 1977, doi: <https://doi.org/10.2307/2529310>.
- [40] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987/11/01/ 1987, doi: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- [41] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197-221, 2011, doi: <https://doi.org/10.1080/00461520.2011.611369>.
- [42] M. T. H. Chi, "Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities," *Topics in Cognitive Science*, vol. 1, no. 1, pp. 73-105, 2009, doi: <https://doi.org/10.1111/j.1756-8765.2008.01005.x>.

Authors' Profiles



January F. Naga is a faculty member at Mindanao State University–Iligan Institute of Technology. Her research interests include information systems, social computing, and health informatics.



Julieta E. Perez is a faculty in the Department of Computer Science at MSU - Iligan Institute of Technology. He is a PhD candidate in Computer Science from De La Salle University - Manila. His research interest revolves around Artificial Intelligence in Education, with a specific focus on the innovative use of Natural Language Processing in building educational applications.

How to cite this paper

January F. Naga, Julieta E. Perez, "Learner Engagement State Typologies in AI Tutoring: A Clustering Analysis of Dialogue Behaviors in Introductory Programming Sessions ", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.18, No.3, pp. 152-166, 2026. DOI:10.5815/ijmeecs.2026.03.10