Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Visualizing Motivational Interviewing: A Two-Study Evaluation of Progress Bars and Behavior Change Features in a Procrastination Conversational Interface

Ben Chen

B.Chen@lboro.ac.uk

Loughborough University https://orcid.org/0009-0005-7488-6515

Nina Dethlefs

Loughborough University

Research Article

Keywords: Conversational AI, Procrastination, Behavior Change, Motivational Interviewing, Human–AI Interaction, Digital Interventions

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11844758/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Graphical Abstract

Visualizing Motivational Interviewing: A Two-Study Evaluation of Progress Bars and Behavior Change Features in a Procrastination Conversational Interface

Ben Chen, Nina Dethlefs

Visualizing Motivational Interviewing: A Two-Study Evaluation of Progress Bars and Behavior Change Features in a Procrastination Conversational Interface

Ben Chen, Nina Dethlefs

[a]*Department of Computer Science, Loughborough University, Loughborough, United Kingdom*

Abstract

Procrastination is a common self-regulation challenge, yet many digital interventions emphasize reflection more than action. We present ProactiMate, an MI (Motivational Interviewing)-based conversational system designed to help users overcome procrastination. Through two studies, we examine how conversational AI can better support motivation and action. Study 1 evaluated a progress bar that visualizes users' advancement through the MI process. Results show that it improved usability, confidence, and positive emotional responses, while fostering a greater sense of progress. Study 2 evaluated an expanded system with features such as progress tracking, micro-interventions, if–then planning, a Pomodoro timer, rewards, and avatar personalization. Findings highlight a persistent gap between reflection and action: although users valued reflective dialogue, they primarily needed concise, actionable guidance to begin tasks. They preferred features that translated reflection into concrete actions. We then discuss design implications for conversational behavior change systems, emphasizing actionability, reduced cognitive load, and meaningful progress feedback.

Keywords: Conversational AI, Procrastination, Behavior Change, Motivational Interviewing, Human–AI Interaction, Digital Interventions

Email addresses: B.Chen@lboro.ac.uk (Ben Chen), N.Dethlefs@lboro.ac.uk (Nina Dethlefs)

1. Introduction

Procrastination is closely associated with self-efficacy, which influences individuals' goal setting, effort, and behavioral outcomes. Prior HCI research has shown that interface features such as progress bars and feedback mechanisms can enhance self-efficacy and support task performance (Jeong et al., 2025). However, most behavior change interventions focus on domains with standardized outcome measures, such as physical activity and digital wellbeing, while areas like procrastination lack consistent evaluation frameworks.

Motivational Interviewing (MI) provides a theory-driven, client-centered approach to behavior change through four processes: Engaging, Focusing, Evoking, and Planning (Miller and Rollnick, 2012). Although MI has been widely applied in human-led interventions, how to effectively operationalize and visualize MI processes within conversational AI systems remains an open research question in HCI.

To address this gap, we developed ProactiMate, a web-based conversational platform that integrates MI principles. We conducted two complementary user studies:

Study 1: A between-subjects experiment evaluating the effect of a dynamic progress bar — representing the three MI stages — on usability, workload, engagement, and procrastination-related behavior. Participants were randomly assigned to a control condition (no progress bar) or an experimental condition (progress bar).

Study 2: A think-aloud study evaluating an extended version of ProactiMate, which added behavior change support tools: progress detection and micro-interventions, avatar personalization, a Pomodoro timer, an if-then planning tool, and reward tracking. This study aimed to uncover how users perceive and interact with these integrated features in a procrastination scenario.

Our findings from Study 1 show that the progress bar significantly improved usability, user confidence, and perceived performance, while moderately increasing perceived time pressure. Findings from Study 2 indicate that participants generally valued the conversational support, motivational structure, and interactive features of the system, particularly for encouraging task initiation and maintaining engagement over time. At the same time, the results reveal an important design tension between reflective guidance and users' preference for concise, actionable instructions. Participants appreciated the supportive conversational interaction but also reported challenges

related to verbosity, cognitive load, and translating reflection into immediate action. These findings highlight the importance of balancing motivational dialogue with clear, action-oriented support in conversational behavior change systems.

Based on these findings, we derive design implications for LLM-powered conversational systems that support behavior change. We show that effective systems should prioritize actionable, step-by-step guidance, minimize cognitive and interactional overhead, and tightly couple reflection with opportunities for immediate action. This work contributes to the design of conversational interfaces that move beyond reflection to more effectively support task execution and sustained engagement.

This paper makes the following contributions:

- **Design of an action-oriented conversational system:** We present a conversational interface that integrates reflective dialogue with lightweight, action-oriented interventions (e.g., adaptive feedback, progress detection, if–then planning, and micro-interventions) to support task engagement and reduce procrastination.
- **Empirical insights from two user studies:** Through a controlled evaluation of a progress indicator (Study 1) and a think-aloud study of the full system (Study 2), we provide evidence on how users interact with and perceive conversational support for task execution.
- **Identification of key barriers to action:** We show that actionability, verbosity, cognitive load, and unclear interface representations are primary challenges in conversational systems designed for behavior change.
- **Design implications for LLM-based behavior change systems:** We derive actionable guidelines for designing effective conversational systems, emphasizing step-by-step action scaffolding, low interaction cost, and tight coupling between reflection and execution.
- **Advancing understanding of reflection–action dynamics:** We contribute to HCI research by highlighting the limitations of reflection-focused conversational approaches and demonstrating the need for systems that actively support the transition from intention to action.

2. Background and Related Work

2.1. Procrastination and Behavior Change

Procrastination is a common self-regulation failure characterized by the voluntary delay of intended actions despite anticipated negative consequences (Thakkar, 2009). Prior research links procrastination to task aversiveness, low self-efficacy, and emotion regulation difficulties (Orpen, 1998). Behavior change support systems (BCSS) and persuasive technologies aim to influence user behavior through mechanisms such as feedback, goal setting, and rewards (Oinas-Kukkonen, 2010). Implementation intentions ("if–then" planning) have been shown to increase the likelihood of action by linking situational cues with specific behaviors (Gollwitzer, 1999). Similarly, micro-interventions—small, actionable prompts—can reduce barriers and support incremental progress (Baumel et al., 2020). However, many existing systems rely on predefined workflows and lack the flexibility to adapt to users' dynamic needs.

2.2. Motivational interviewing (MI)

Motivational interviewing (MI) was created to assist individuals in resolving ambivalence and making a commitment to change (Hettema et al., 2005). The practice of motivational interviewing is grounded in a client-centered approach (Rollnick and Miller, 1995). There are four processes for MI, they are Engaging, Focusing, Evoking and Planning. Engaging refers to build rapport and trust to establish a collaborative relationship. Focusing means narrowing the conversation to specific goals or areas of change. Evoking stands for eliciting the client's own reasons for change and resolving ambivalence. And Planning symbolizes developing a concrete plan for action and change (Miller and Rollnick, 2012).

The Motivational Interviewing Skill Code (MISC) was designed to evaluate the quality of MI (Miller et al., 2003). It categorizes counselor behaviors (such as asking open questions, providing reflections, and offering affirmations) and client language (like change talk or sustain talk) to evaluate engagement, ambivalence, and commitment to change (Miller et al., 2003). Refer to Table 2.2 for a summary of the classification categories.

2.3. Conversational Agents for Behavior Change

Recent research shows that conversational conversational interface can support behavior change by integrating psychological frameworks such as

Table 1: Description of sentence codes based on MI in three levels. This table categorizes sentences according to MI principles using a three-level coding scheme. The Valence level classifies sentences as Change Talk (positive, +), Sustain Talk (negative, -), or Follow/Neutral (FN) based on their motivational orientation. The Content Label level further categorizes sentences into Reason (R), Taking Steps (TS), and Commitment (C), representing different aspects of motivation and action. Finally, the Reason Sublabel level provides finer distinctions within the "Reason" category, identifying whether a statement reflects ability (a), need (n), desire (d), or general justifications and incentives.

Level	Description
Valence	Change Talk(+), Sustain Talk(-), Follow/Neutral(FN)
Content Label	Reason(R), Taking Steps(TS), Commitment(C)
Reason Sublabel	ability(a), need(n), desire(d), general justifications, incentives, or justifications(general)

Motivational Interviewing (MI) and the COM-B model (Miller and Rollnick, 2012; Hegde et al., 2024). MI-based systems emphasize self-reflection, intrinsic motivation, and empathetic dialogue (Kumar, 2024; Ahmed et al., 2022; Meyer and Elsweiler), whereas COM-B-based conversational interface provide structured and measurable interventions targeting behavioral determinants (Chang et al., 2024; van Baal et al., 2022). Other studies incorporate personalization, wearable data integration, and optimized prompting techniques to improve engagement and intervention effectiveness (Jörke et al., 2024; Nepal et al., 2024). Existing conversational interface interventions vary considerably in terms of theoretical foundations, interaction styles, platforms, and evaluation methods, demonstrating both the flexibility and complexity of designing conversational systems for sustainable behavior change and procrastination management (Bhattacharjee et al., 2024).

2.4. Progress Indicators in Conversational Contexts

Progress indicators and similar interface elements are widely used in graphical interfaces to enhance perceived control and motivation (Harrison et al., 2007).

More recently, researchers have begun exploring how progress indicators can be integrated into conversational systems. Daeun et al. developed an interface aimed at enhancing self-efficacy during task performance, consist-

ing of two main components: a Progress Bar and a Subtask Marker (Jeong et al., 2025). Together, these features provide a visual representation of overall progress and deliver feedback on subtask completion. Results from the user study show that the progress bar enhances users' perceived ability to complete tasks, making it particularly beneficial in multi-turn conversational settings where maintaining motivation is critical (Jeong et al., 2025).

However, the role of progress indicators in motivational interviewing (MI) contexts remains underexplored. MI-based conversational systems are primarily designed to support reflection, ambivalence resolution, and intrinsic motivation rather than task completion or linear progression (Miller and Rollnick, 2012). As a result, introducing explicit progress representations in MI contexts may influence the conversational dynamic. While such indicators could reinforce a sense of momentum or accomplishment, they may also risk imposing a sense of linearity or external pressure that conflicts with the non-directive and autonomy-supportive principles of MI (Deci and Ryan, 2000; Miller and Rollnick, 2012). Prior work emphasizes that supporting autonomy and intrinsic motivation is critical in behavior change interventions, and overly directive or externally imposed structures may undermine engagement or reduce perceived authenticity.

Given these challenges, there is a need to better understand whether and how progress indicators can be meaningfully integrated into MI-based conversational systems. In particular, it remains unclear how such representations affect user experience, perceived progress, and motivation when the interaction is not strictly task-oriented. This gap motivates the present work, which examines the effectiveness and limitations of progress indicators within conversational settings that combine reflective dialogue with goal-directed support.

2.5. Summary

Motivational Interviewing (MI) emphasizes reflection, autonomy, and intrinsic motivation, yet provides limited support for translating insight into action. Conversational agents extend these approaches by enabling interactive and scalable interventions, though they still struggle to bridge the gap between reflection and execution. While progress indicators have been shown to enhance motivation and self-efficacy in graphical interfaces and emerging conversational systems, their role in MI-based, open-ended interactions remains unclear. This tension between structured progress representation and

reflective dialogue highlights the need to better understand how such feedback can support action and engagement in conversational settings.

3. Interface Design

Building on insights from the preceding literature review, we developed an interactive platform aimed at supporting procrastination avoidance, guided by the conceptual framework established in prior work (Chen and Dethlefs, 2025). The interactive platform, ProactiMate, was developed as a web-based conversational interface designed to facilitate user engagement in procrastination mitigation. The interface was implemented using React.js, a component-based front-end framework that supports dynamic and responsive user experiences. The system enables seamless, real-time dialogue between the user and the chatbot through asynchronous communication with a Flask-based backend via RESTful API calls.

The interface begins with an initial system message, introducing the website as a motivational assistant whose purpose is to guide users in understanding and addressing procrastination. The chat interface adopts a turn-based conversational flow, displaying user and assistant messages in a structured and visually distinct format to maintain clarity and engagement. As discussed in prior work (Chen and Dethlefs, 2025), ProactiMate structures the conversational flow into three distinct stages, each aligning with the four core processes of Motivational Interviewing (MI): engaging, focusing, evoking, and planning (Clifford and Curtis, 2016). The dialogue is designed as three conversational turns for Stage 1, three turns for Stage 2 and one turn for Stage 3, providing a systematic progression from rapport building to goal formulation and action planning.

The progress bar component was implemented to visualize users' advancement through the conversational stages of the system. The progress bar is placed at the top of the conversational interface window to provide a clear and immediate overview of progression. The component comprises two main parts: a series of textual labels indicating the conversation stages (Engaging & Focusing, Evoking, and Planning), and a dynamic horizontal bar that fills progressively as the user transitions between stages. From a behavioral design perspective, the progress bar serves as a self-regulatory cue, offering users immediate feedback on their advancement through the MI-guided conversation flow. The initial interface for facilitating the progress bar is shown in Figure 1. After the conversation reaches the second stage, the Evoking

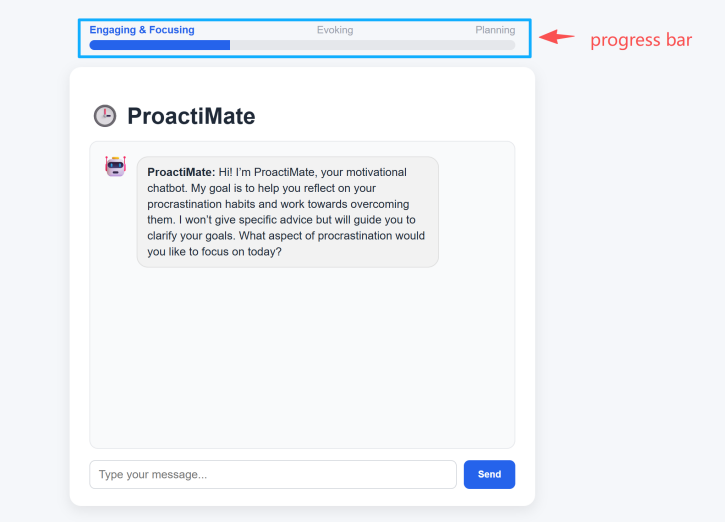

Figure 1: Initial interface

stage, the interface is shown in Figure 2. When the conversation reaches its final stage, further interaction is prevented by disabling the text input box and send button.

4. Study 1: Understanding the Role of Progress Indicators in Conversational Task Support

To evaluate the effectiveness of the progress bar, we formulated the following three hypotheses.

- Participants using the interface with a progress bar will report higher perceived clarity of conversation flow than those using the interface without it.
- Participants using the progress bar will show higher motivation and engagement scores.
- The progress bar will lead to shorter perceived task time and lower cognitive effort, reflecting reduced uncertainty.

4.1. Participants

A total of 24 participants were recruited for the study (12 male, 12 female), with ages ranging from 26 to 56 years (M = 34.42). Participants were

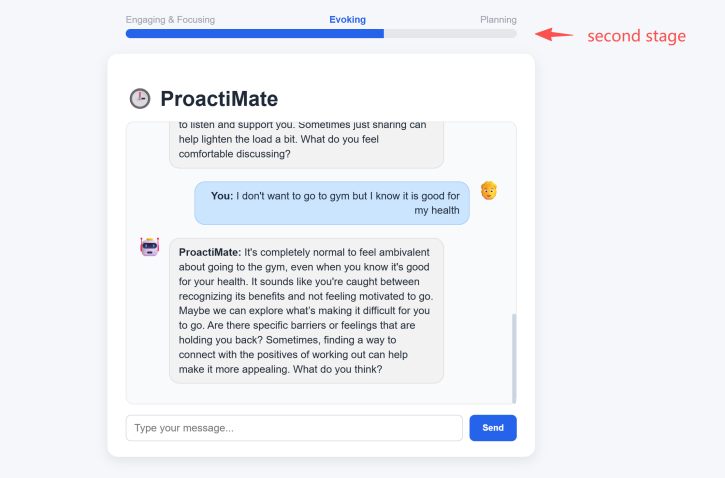

Figure 2: Second stage

evenly distributed across gender and represented a broad adult age range, providing diverse perspectives on the interaction with the system. Participants were recruited through online advertisements and voluntarily took part in the study. Informed consent was obtained from all participants prior to participation. Participants were assigned to either the control group or the experimental group while maintaining gender balance across conditions, with 12 participants in each group.

4.2. Procedure

Participants were first welcomed by the researcher, who provided an overview of the study and explained its purpose. It is to evaluate users' interaction experience with the conversational interface ProactiMate.

Following consent, participants received a brief training session introducing the conversational interface and explaining how to engage in the conversation. This orientation ensured participants were familiar with the system's layout, input method, and overall interaction flow.

During the task phase, participants interacted with ProactiMate and completed a short, guided dialogue designed to simulate a procrastination-reflection scenario. This scenario was self-generated by the participants. Participants were asked to reflect on and identify a task they had been procrastinating on most frequently in their recent daily life. We deployed between-subjects experiment with two interface conditions. Condition A was served

as control group, in which participants were presented with the interface without a progress bar on the top. Condition B was served as experimental group and participants were presented with a progress bar on the top showing the three MI stages dynamically. The interface interaction comprised three distinct stages: Engaging and Focusing, Evoking, and Planning, which are the core processes of MI theory. Accordingly, participants were evenly distributed between the control group (n = 12) and the experimental group (n = 12).

After completing the conversational interface interaction, participants proceeded to the post-interaction phase, which involved completing several standardized questionnaires to assess their experience. These included the System Usability Scale (SUS) to evaluate usability, the NASA Task Load Index (NASA-TLX) to assess perceived workload, and the User Engagement Scale (UES) to measure engagement and motivation. All questionnaire items were measured using rating scales to enable quantitative analysis.

Finally, participants took part in a semi-structured interview aimed at gaining qualitative insights into their experience. The interview was conducted and recorded via Microsoft Teams with the participants' informed consent, and the transcript was subsequently used for qualitative analysis.

4.3. Results

A comparative analysis was conducted to examine differences in user experience between the control group and the experimental group. Overall, participants in the experimental group reported more positive experiences among engagement, usability and perceived performance compared to those in the control group.

In terms of usability, the experimental group reported significantly greater ease of use, learnability, and intention for future use than the control group. Please refer to Table 2 for questionnaire design and Figure 3 for the grouped bar chart of the SUS. Specifically, for the item "I think that I would like to use this system frequently ("Intention for Future Use" item)," the experimental group showed a higher mean score (M = 3.08) than the control group (M = 2.33), with the pairwise comparison indicating a statistically significant difference (p = .04). Similarly, for "Learnability", the experimental group also reported higher ratings (M = 3.00) compared to the control group (M = 2.58), with this difference reaching statistical significance (p = .01). Participants in this condition perceived the system as easier to learn and expressed stronger willingness to continue using it, whereas participants in the control

Table 2: System Usability Scale (SUS) questionnaire items used in the study. Responses were recorded on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Item	**Statement**
SUS1	I think that I would like to use this system frequently.
SUS2	I found the system unnecessarily complex.
SUS3	I thought the system was easy to use.
SUS4	I think that I would need the support of a technical person to be able to use this system.
SUS5	I found the various functions in this system were well integrated.
SUS6	I thought there was too much inconsistency in this system.
SUS7	I would imagine that most people would learn to use this system very quickly.
SUS8	I found the system very cumbersome to use.
SUS9	I felt very confident using the system.
SUS10	I needed to learn a lot of things before I could get going with this system.

group reported comparatively lower usability perceptions. Besides, participants in the experimental group reported a high level of confidence when using the system (M = 4.25), which was notably higher than the confidence reported by the control group (M = 3.75). Overall, the use of progress bar demonstrated significantly higher usability perceptions, including ease of use, learnability, intention for future use, and confidence, compared to the control group without progress bar, indicating a more positive overall user experience with the system. In terms of user engagement, please refer to Table 3 for questionnaire design and Figure 4 for the grouped bar chart of the UES. For the item "Focused Attention" participants in the experimental group reported a lower mean score (M = 2.67) compared to the control group (M = 2.97). This suggests that the experimental condition may have introduced elements that slightly disrupted user concentration or immersion relative to the baseline control. For the item "Reward / Perceived Value" participants in the experimental group reported higher ratings (M = 3.58) than those in the control group (M = 3.42), suggesting the experimental condition offered slightly greater perceived utility, benefit, or intrinsic motivation. Overall, the progress bar intervention appears to involve a trade-off: while it slightly reduces users' focused attention, it enhances perceived reward and value. This

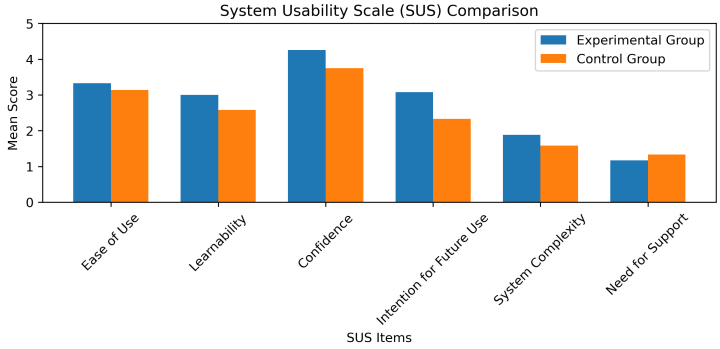

Figure 3: System Usability Scale (SUS) Comparison

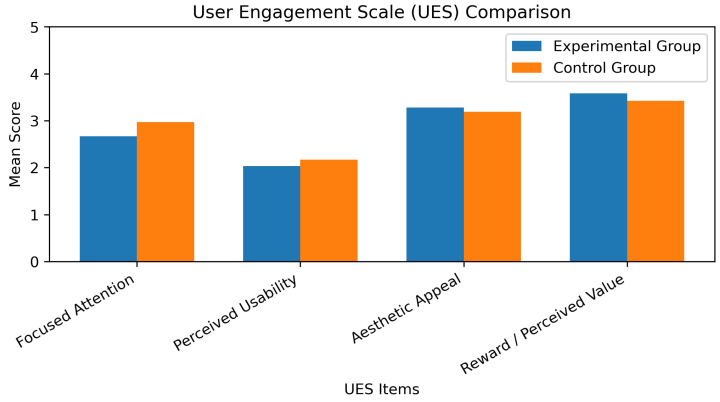

Figure 4: User Engagement Scale (UES) Comparison

suggests that the progress bar increases users' sense of usefulness or motivation, but may require refinement to better support sustained attention and immersion. With respect to workload, participants in the experimental group reported a significantly higher sense of time pressure than those in the control group. Table 4 below shows questionnaire design and Figure 5 below provides detailed information on the NASA radar chart. Specifically, for the item "How hurried or rushed was the pace of the task?" ("Temporal demand" on the radar chart), the experimental group exhibited a higher mean score ($M = 3.25$) compared to the control group ($M = 2.08$), with this difference reaching statistical significance ($p = .04$). This finding suggests that the presence of the progress bar may have increased participants' awareness of task pacing, potentially encouraging more efficient task progression dur-

Table 3: User Engagement Scale (UES) items used in the study. Responses were recorded on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Dimension	**Item**
Focused Attention	I lost myself in this experience.
Focused Attention	The time I spent using the chatbot just slipped away.
Focused Attention	I was absorbed in this experience.
Perceived Usability	I felt frustrated while using this chatbot.
Perceived Usability	I found this chatbot confusing to use.
Perceived Usability	Using this chatbot was taxing.
Aesthetic Appeal	This chatbot was attractive.
Aesthetic Appeal	This chatbot was aesthetically appealing.
Aesthetic Appeal	The chatbot appealed to my visual senses.
Reward	Using this chatbot was worthwhile.
Reward	My experience was rewarding.
Reward	I felt interested in this experience.

ing the interaction. Besides, participants in the experimental group reported slightly higher levels of success in accomplishing the assigned tasks compared to the control group. Specifically, for the item "How successful were you in accomplishing what you were asked to do?" ("Performance"), the experimental group reported a higher mean score (M = 5.25) than the control group (M = 4.92). This difference suggests that the progress bar may have supported more effective task completion, although the magnitude of the difference was modest. Besides, participants in the experimental group reported higher levels of effort required to accomplish their level of performance compared to the control group. Specifically, for the item "How hard did you have to work to accomplish your level of performance?" ("Effort"), the experimental group

Table 4: NASA-TLX questionnaire items. All items were rated on a 7-point Likert scale (1 = Very Low, 7 = Very High).

Category	**Question**
Mental Demand	How mentally demanding was the task?
Physical Demand	How physically demanding was the task?
Temporal Demand	How hurried or rushed was the pace of the task?
Performance	How successful were you in accomplishing what you were asked to do?
Effort	How hard did you have to work to accomplish your level of performance?
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?

reported a higher mean score (M = 3.83) than the control group (M = 2.92). Apart from that, participants in the experimental group reported lower levels of insecurity, discouragement, irritation, stress, and annoyance compared to the control group. Specifically, for the item "How insecure, discouraged, irritated, stressed, and annoyed were you?" ("Frustration"), the experimental group exhibited a lower mean score (M = 2.75) than the control group (M = 3.17). This result suggests that the progress bar may have mitigated negative emotional responses during task execution, contributing to a more positive overall user experience. Overall, the progress bar intervention led to a more positive user experience compared to the control condition. It significantly improved perceived usability, including ease of use, learnability, intention for future use, and confidence. While the progress bar slightly reduced focused attention and increased perceived time pressure and effort, it enhanced perceived reward, task performance, and reduced frustration. These findings suggest that the progress bar effectively supports motivation and task completion, though its design may be further optimized to minimize attentional disruption and workload.

A comparative thematic analysis was performed on data from the experimental and control groups based on semi-structured interview, and the identified themes are presented in Table 5.

Theme 1: Reflection and Awareness of Procrastination

Participants in both conditions reported that the conversational interface facilitated self-reflection on procrastination and task-related motivation. In the experimental group, users described the interaction as prompting a grad-

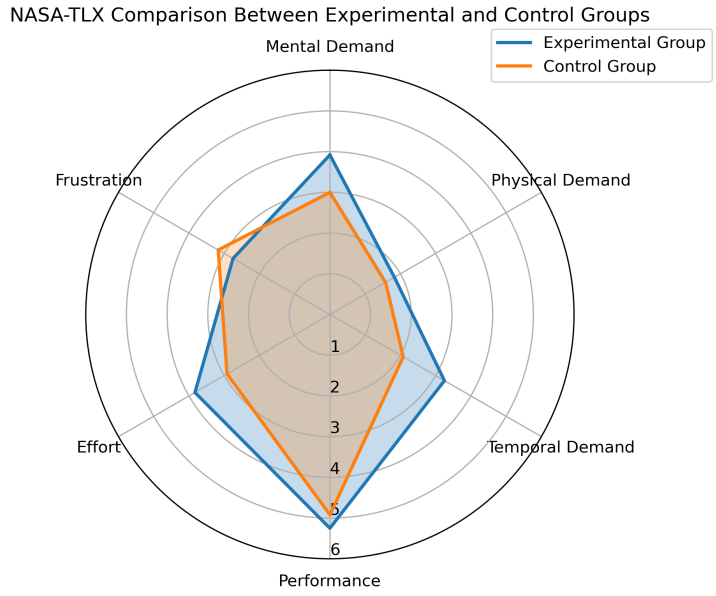

Figure 5: NASA-TLX Comparison Between Experimental and Control Groups

ual shift in mindset, encouraging them to reconsider priorities and remain aware of pending tasks, even without concrete solutions. This reflective process aligned with principles of Motivational Interviewing, functioning primarily as a trigger for awareness rather than immediate behavior change. Similarly, participants in the control group noted that the conversational interface helped surface underlying barriers—such as task ambiguity, avoidance, and emotional discomfort—making reflection valuable, particularly during early-stage problem exploration.

Theme 2: Motivation and Action Readiness

Participants in the experimental group generally reported increased motivation and intention to act, often describing a boost in readiness toward the end of the conversation. However, this effect was primarily short-term, with limited support for sustained behavior change. In contrast, responses in the control group were more varied. While some participants felt encouraged, others reported little impact, emphasizing that motivation alone was insufficient to overcome procrastination. Instead, action readiness was seen as dependent on task clarity and concrete guidance.

Theme 3: Perceived Conversational Quality

Conversational quality played a significant role in user engagement across

Table 5: Summary of Thematic Findings by Study Condition

Theme	**Experimental Group**	**Control Group**
Reflection	Encouraged self-awareness; acted as reflective trigger rather than driving action	Increased awareness of barriers; valuable for early-stage exploration
Motivation	Boosted short-term motivation; limited support for sustained action	Mixed effects; action depends more on clarity than encouragement
Conversation Quality	Engagement improved with clear, varied responses; reduced by repetition and verbosity	Preferred personalized, responsive dialogue; disliked long or scripted responses
Progress Bar	Sometimes useful as structural cue; often unnoticed or weakly linked to progress	Mixed views; helpful for structure but unclear in open-ended interaction
Usability	Need for clearer guidance and actionable outputs	Issues with flow, verbosity, and lack of step-by-step support

both conditions. In the experimental group, engagement improved when responses encouraged reflection without being repetitive, while overly dense or literal responses reduced effectiveness. In the control group, participants valued interactions that felt responsive, personalized, and dialogic. However, long responses, repetitive suggestions, and perceived misunderstandings led to disengagement, making the interaction feel scripted or artificial.

For example, when a participant reported avoiding a dissertation chapter, the chatbot responded: "What makes this task feel difficult to start right now?" followed by "What would be one small step you could take in the next 10 minutes?" Participants described these prompts as helpful because they encouraged reflection while leading toward action. In contrast, responses containing multiple reflective questions or lengthy explanations were often perceived as overwhelming. One participant noted that repeated recommendations to "just start with a small step" became less useful over time when not adapted to their specific situation.

Theme 4: Progress Bar

Participants in the experimental group expressed mixed perceptions of the progress bar. When noticed, it served as a useful structural cue, helping users understand conversation flow and anticipate completion. However,

many participants did not notice it initially or only became aware of it later, limiting its overall impact. In the control group, opinions were similarly mixed. Some participants found progress indicators helpful for structure and orientation from hypothetical questions, while others viewed them as incompatible with open-ended conversations. Overall, effectiveness depended on whether the progress bar was clearly linked to task structure and goals.

Theme 5: Usability Challenges and Design Expectations

Participants in both conditions identified several usability challenges. In the experimental group, users reported confusion about next steps, unclear conversation endings, and a desire for more explicit stage explanations and actionable outputs such as task breakdowns. In the control group, participants highlighted issues related to conversational flow, excessive text length, multiple questions per turn, and limited context awareness. Across both groups, there was a clear expectation for more adaptive, structured, and step-by-step guidance to support task execution.

In conclusion, among both conditions, the conversational interface primarily supported reflection and short-term motivation, while sustained action depended on task clarity and conversational structure. The progress bar offered additional orientation and control when noticed, but its impact was limited by low salience and the open-ended interaction design. Control group participants viewed the progress bar as potentially helpful for structure and orientation, but only when paired with clearly defined tasks and stages. Overall, the findings highlight the need to integrate reflective conversational support with clearer structural and action-oriented elements in conversational AI design.

4.4. Discussion

From quantitative analysis, results indicate that the progress bar positively influenced overall user experience by improving usability, perceived performance, and emotional responses. Participants found the system easier to learn, felt more confident using it, and expressed greater willingness to continue use when a progress bar was present. Although the progress bar slightly reduced focused attention and increased perceived time pressure and effort, it also enhanced perceived reward and reduced frustration. These findings suggest that progress bars can effectively motivate users and support task completion, but their design should be carefully balanced to minimize attentional disruption and unnecessary workload.

From thematic analysis, the conversational interface primarily supported reflection and awareness of procrastination across both groups, aligning with Motivational Interviewing principles. Participants valued the opportunity to reconsider their mindset and underlying barriers, though this reflection did not consistently translate into immediate action.

The experimental group reported more consistent short-term motivation, suggesting that the progress bar and structured interaction may have helped create a sense of momentum and completion.

Conversational quality strongly influenced engagement in both conditions. Participants preferred concise, adaptive, and responsive interactions, while repetitive or overly dense responses reduced engagement. Usability feedback further highlighted the need for clearer conversational flow, better closure cues, and more concrete outputs such as task breakdowns. Overall, the findings suggest that combining reflective dialogue with clearer structure, actionable guidance, and more salient progress cues may enhance user engagement and behavioral impact.

5. Interface Extension

Building on findings from our prior usability study, we refined and expanded the ProactiMate interface to better support the transition from reflection to action.

The new interface adopts a two-panel layout. The left panel focuses on the conversational interaction and motivational guidance, while the right panel provides supporting behavior-change tools such as planning and reward tracking.

5.1. Engagement Feedback

The system provides adaptive engagement feedback based on behavioral signals during the conversation. Engagement is estimated using two indicators: the average length of user messages and the time taken for users to respond to the conversational interface. Prior research in human–computer interaction suggests that conversational engagement can be inferred from behavioral interaction patterns such as response latency and message length, which reflect user attention, cognitive effort, and involvement in dialogue (Oertel et al., 2020; Martinengo et al., 2022).

Based on these indicators, the interface displays contextual feedback messages indicating high, medium, or low engagement levels. For example, highly

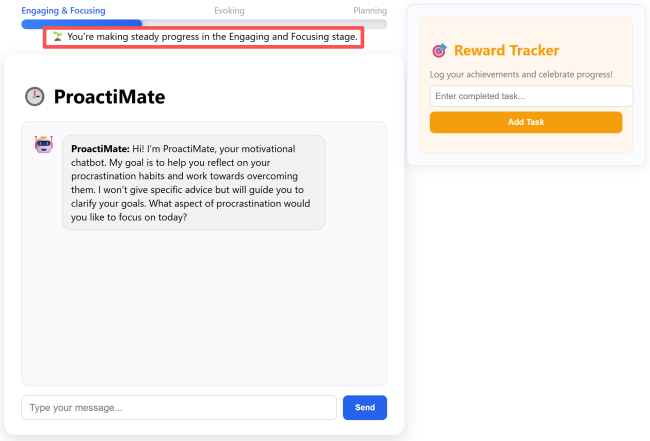

Figure 6: The system provides adaptive engagement feedback during the conversation based on behavioral signals, specifically user response latency and message length. These indicators are used to infer engagement levels (high, medium, low), which trigger corresponding feedback messages—ranging from reinforcement for highly engaged users to gentle encouragement for lower engagement—to support motivation and sustained interaction.

engaged users receive reinforcement messages acknowledging their momentum, while users with lower engagement receive gentle encouragement emphasizing that small steps are valuable. Providing adaptive feedback aligns with principles from persuasive technology and behavior change systems, where timely feedback can strengthen user motivation and support sustained engagement with the intervention (Fogg, 2002; Oinas-Kukkonen, 2010).
The interface is illustrated in Figure 6.

5.2. Progress Detection and Action Suggestions

The interface includes a lightweight progress detection mechanism that analyzes user messages to identify potential progress cues. User responses are automatically classified using the GPT-3.5-turbo model, which acts as a zero-shot text classifier. For each user message, the model evaluates the content and assigns it to one of four predefined categories:

- Completion: the user has finished a task
- Partial progress: the user has made some progress
- Effort: the user attempted the task but progress is unclear

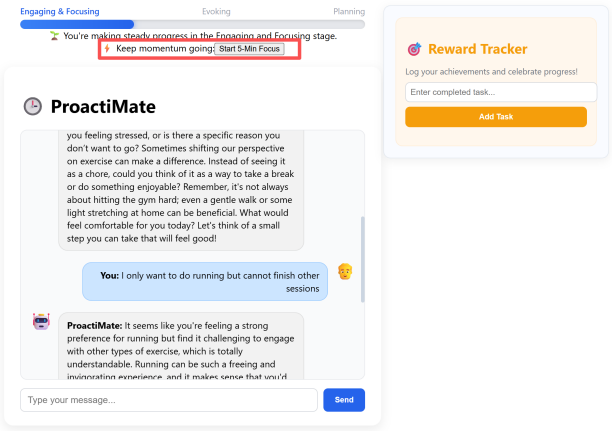

Figure 7: The interface incorporates a lightweight progress detection mechanism that classifies user responses into four categories—completion, partial progress, effort, and no progress—based on conversational cues. These classifications trigger contextual micro-interventions, such as prompting short focus sessions or suggesting new work sprints, to support ongoing task engagement. This adaptive approach leverages behavior change principles by providing timely, actionable feedback that lowers barriers to action and encourages incremental progress.

- None: no progress is indicated

Based on the detected category, the interface provides contextual micro-interventions. For example, users who indicate partial progress may be encouraged to start a short five-minute focus session, while users reporting completion may be prompted to begin a new short work sprint.

Providing adaptive suggestions based on user progress aligns with principles from behavior change support systems and persuasive technology, where timely feedback and tailored interventions can reinforce motivation and support self-regulation (Fogg, 2002). In particular, small actionable prompts such as short focus sessions function as micro-interventions that lower the barrier to action and encourage incremental progress toward larger goals (Baumel et al., 2020). This approach is also consistent with research on habit formation and self-regulation, which suggests that breaking tasks into small, achievable steps can increase perceived self-efficacy and reduce procrastination-related avoidance behaviors (Sullivan, 2020). The interface is illustrated in Figure 7.

5.3. Avatar Personalization

The chatbot avatar is displayed alongside assistant messages and can be personalized through an emoji-based avatar picker. Users can select from several predefined categories representing different conversational personas, including AI, Coach, Motivator, Calm, and Friendly. Each category contains a set of emoji-based visual representations that reflect different tones or styles of interaction. By allowing users to select an avatar that matches their preferred interaction style, the system aims to foster a stronger sense of social presence and personalization within the conversation.

The design of avatar personalization is grounded in social presence theory, which posits that the perceived awareness of another social entity in a mediated environment enhances engagement, trust, and relational closeness (Fortunati and Manganelli, 2008). By providing avatars that convey different conversational personas, the system can simulate aspects of social presence, making interactions with the AI feel more human-like and socially responsive (Bente et al., 2008). Furthermore, research in human-computer interaction (HCI) suggests that personalized visual cues, such as avatars, can increase user satisfaction, motivation, and perceived control over the system (Schroeder, 2002).

Similarly, users can choose their own avatar using a simplified selection interface. The selected avatar appears next to all user messages in the conversation, creating a consistent visual identity throughout the interaction. This consistency supports identity expression and continuity, which are important for fostering engagement and a sense of presence in conversational systems (Munnukka et al., 2022).

Avatar selections are stored locally in the browser to ensure that user preferences persist across sessions. This lightweight personalization mechanism is intended to make the interaction feel more engaging and conversational while maintaining a simple and accessible interface design. The interface is illustrated in Figure 8.

5.4. Pomodoro Timer

During the Planning stage, the interface introduces a built-in Pomodoro timer to support focused work sessions. The timer allows users to start, pause, or reset a countdown session, with the default duration set to 25 minutes. When a session ends, the conversational interface prompts the user to reflect on what they accomplished during the focus period. This fea-

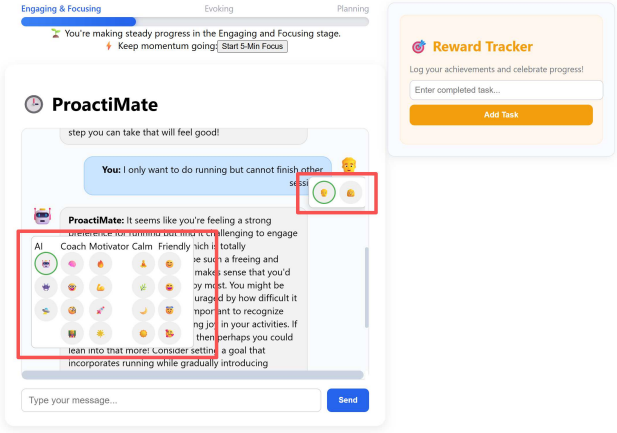

Figure 8: Avatar personalization allows users to select emoji-based personas for both the chatbot and themselves, creating a consistent visual identity that enhances social presence and engagement.

ture integrates structured time management directly into the conversational workflow.

The design of the Pomodoro timer is grounded in time management and self-regulation theories, which suggest that structured intervals of focused work, coupled with brief breaks, improve attention, productivity, and cognitive performance (Almalki et al., 2020; Zimmerman, 2000). The Pomodoro technique itself utilizes principles of goal-setting and behavioral reinforcement, segmenting tasks into manageable intervals that reduce cognitive load and procrastination tendencies while providing immediate feedback on progress (Almalki et al., 2020).

By combining structured time intervals, progress reflection, and conversational guidance, this feature aims to support sustained attention, reduce task avoidance, and reinforce productive work habits within the workflow of the system. The interface is illustrated in Figure 9.

5.5. If–Then Planning Tool

To support implementation planning, the interface includes an If–Then planner that enables users to formulate implementation intentions during the Planning stage. Users can specify a trigger condition ("If...") and a corresponding action ("Then..."). For example, a user might define: "If I finish lunch, then I will spend 10 minutes working on my dissertation."

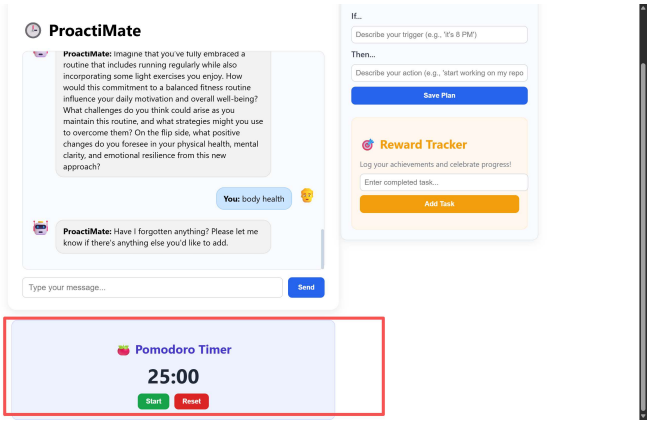

Figure 9: The interface integrates a Pomodoro timer within the planning stage to support focused work sessions, combining structured time intervals with conversational prompts for reflection and sustained task engagement.

This planning structure helps users translate abstract goals into concrete behavioral plans, supporting the transition from intention to action.

The design of this feature is grounded in the concept of implementation intentions, a well-established construct in social and health psychology (Gollwitzer, 1999). Implementation intentions are structured as conditional "if–then" plans that link anticipated situational cues with goal-directed behaviors. By predefining when, where, and how to act, these plans facilitate automatic action initiation when the specified cue is encountered, thereby reducing reliance on effortful decision-making at the moment of execution (Gollwitzer and Sheeran, 2006).

Embedding this planning mechanism within a conversational interface also provides interactive scaffolding, guiding users in constructing effective plans and reinforcing engagement through dialogue. Overall, the If–Then planner employs evidence-based behavior change techniques to support the translation of intentions into sustained action. The interface is illustrated in Figure 10.

5.6. Reward Tracking

The interface also includes a reward tracker where users can record completed tasks. Each recorded achievement is added to a visible list, allowing users to track their progress over time. When users accumulate 5 completed tasks, the system displays a visual badge acknowledging their achievements.

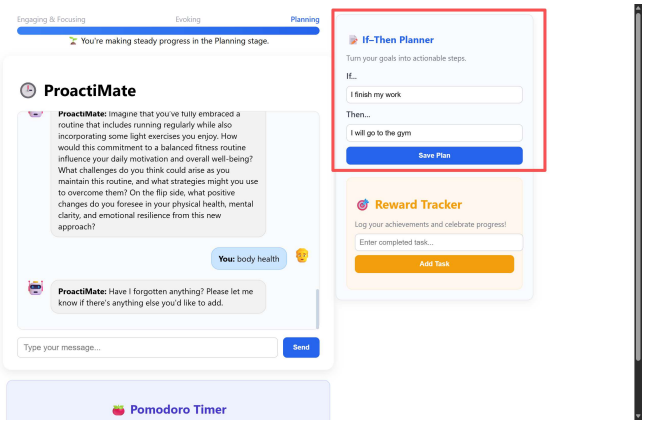

Figure 10: The interface includes an If–Then planner that helps users create implementation intentions by linking situational cues to specific actions, supporting the transition from planning to action.

This mechanism provides positive reinforcement and aims to strengthen motivation and self-efficacy.

The design of this feature is grounded in principles of operant conditioning, which suggest that behaviors followed by positive reinforcement are more likely to be repeated (Skinner, 1965). By rewarding task completion with visible progress indicators and achievement badges, the system reinforces productive behaviors and encourages continued engagement. Such reward-based mechanisms are widely used in digital interventions to promote habit formation and sustained behavior change.

In addition, the reward tracker draws on self-efficacy theory (Locke, 1997), which emphasizes the role of perceived competence in motivating behavior. Recording completed tasks provides users with tangible evidence of their progress, thereby enhancing their belief in their ability to successfully perform goal-directed actions. This aligns with research showing that progress feedback and mastery experiences are key drivers of increased self-efficacy and persistence.

The use of badges and accumulated achievements also reflects concepts from gamification, where game-like elements such as points, rewards, and visual milestones are used to increase user engagement and motivation (Deterding et al., 2011). These elements can foster a sense of accomplishment and provide immediate feedback, which is particularly effective in maintaining user interest over time. The interface is illustrated in Figure 11.

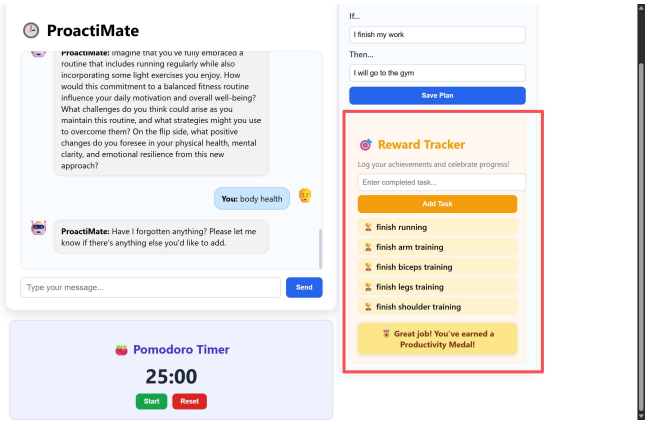

Figure 11: The interface includes a gamified reward tracker that records completed tasks and provides visual badges as positive reinforcement, supporting motivation, self-efficacy, and sustained engagement.

6. Study 2: Think-Aloud Evaluation of the Expanded ProactiMate Interface

We employed a think-aloud protocol (TAP) to evaluate user interaction with ProactiMate. Think-aloud methods are widely used in Human-Computer Interaction to capture users' cognitive processes during task execution, enabling insights into usability, perception, and decision-making (Jääskeläinen, 2010).

6.1. Procedure

Participants (different from Study 1) engaged in a 40–50 minute moderated session, during which they interacted with the system while continuously verbalizing their thoughts. They were instructed:

" Please say out loud what you are thinking — what you expect to happen, what confuses you, what feels helpful, and why you choose certain actions. There are no right or wrong answers; we are evaluating the system, not you."

The study was structured into eight sequential tasks, each targeting specific components of the interface and underlying behavior change mechanisms.

6.2. Tasks

Task 1: First Impression & Initial Interaction

Participants initiated a conversation with the conversational interface and explored the interface freely. This task focused on initial usability, discoverability, and first impressions.

Task 2: Sustained Conversation

Participants continued interacting with the conversational interface about their procrastination behaviors. This task examined perceived empathy, conversational naturalness, and understanding of MI-style prompts.

Task 3: Progress Awareness

Participants were asked to interpret the progress bar representing conversational stages. This assessed comprehension of system feedback and its influence on user engagement.

Task 4: Avatar Customization

Participants explored options to personalize avatars. This task evaluated discoverability and perceived value of personalization features.

Task 5: Actionable Suggestions

Participants triggered system-generated action buttons by reporting progress. This examined responsiveness, perceived usefulness of suggestions, and user expectations of system behavior.

Task 6: Pomodoro Timer Interaction

Participants interacted with a timer introduced during the planning stage. This assessed understanding of time-management support and ease of use.

Task 7: If–Then Planning Tool

Participants created implementation intentions using an if–then planner. This task evaluated comprehension of structured planning and its perceived usefulness for behavior change.

Task 8: Reward Tracking System

Participants used a reward tracker to log completed tasks and observe gamification elements. This examined motivational impact and engagement with progress reinforcement mechanisms.

6.3. Data Collection

We collected both qualitative and quantitative data:

- Think-aloud verbalizations, capturing real-time cognitive processes
- Behavioral observations, including task completion, errors, and help requests

- Interaction logs, such as clicks and time to action
- Affective responses, including positive and negative utterances
- Post-task reflections, focusing on usability, conversational interface quality, and perceived behavioral impact

6.4. Post-Interaction Interview

Following task completion, participants engaged in a semi-structured interview exploring:

- Usability (ease of use, confusion points)
- Conversational interface quality (perceived supportiveness, conversational realism)
- Behavioral impact (likelihood of real-world use, influence on procrastination)
- Design improvements (feature-level feedback and suggestions)

6.5. Analysis Approach

Think-aloud data were analyzed using thematic analysis, focusing on usability issues, perceptions of conversational quality, and engagement with behavior change features. Quantitative metrics were used to contextualize qualitative findings and identify patterns across participants.

6.6. Participants

We recruited five participants to take part in a think-aloud protocol study, a qualitative method commonly used to capture users' cognitive processes while interacting with a system. Each participant was asked to verbalize their thoughts, decisions, and reasoning in real time as they completed the assigned tasks.

The study was conducted remotely via Microsoft Teams to allow for flexible participation and screen sharing. All sessions were audio- and video-recorded with participants' consent. The recordings were subsequently transcribed to support detailed qualitative analysis. These transcripts were then used to identify patterns in user behavior, reasoning strategies, and potential usability issues.

6.7. Findings

We conducted a full-pass coding of the think-aloud transcripts, identifying recurring instances of each code across all participants (N = 5). The unit of analysis was a meaningful utterance, and multiple codes could be applied per segment.

A total of 78 coded instances were identified and the code frequency distribution is displayed in Table 6.

Table 6: Code Frequency Distribution

Code	Label	Frequency	Participants
C1	Perceived Verbosity	9	5/5
C2	Preference for Conciseness	8	5/5
C3	Planning–Execution Gap	10	5/5
C4	Step-by-Step Guidance	7	4/5
C5	Iterative Interaction	5	3/5
C6	Interaction Overhead	8	4/5
C7	Feature Burden	6	4/5
C8	Conceptual Confusion	5	3/5
C9	Progress Bar Disconnect	6	4/5
C10	System Assumption Frustration	4	3/5
C11	Chatbot as Primary Value	7	5/5
C12	Feature Disconnection	6	4/5
C13	Time Pressure Interpretation	3	2/5
C14	Need for Clarity & Motivation	7	5/5
C15	Task-Oriented Mindset	9	5/5
C16	Lack of Practical Usefulness	3	3/5

The codes are then organized thematically, as presented in Table 7.

We then generated bar chart, heatmap and stacked bar chart for codes distribution in Figure 12, Figure 13 and Figure 14. We conducted a thematic analysis based on Table 6 and Table 7 to understand how users interact with the conversational interface system and how it supports (or fails to support) task engagement and procrastination management. Four key themes emerged: (1) preference for conversational simplicity, (2) breakdown between reflection and action, (3) interface comprehension challenges, and (4) mixed effectiveness of motivational features.

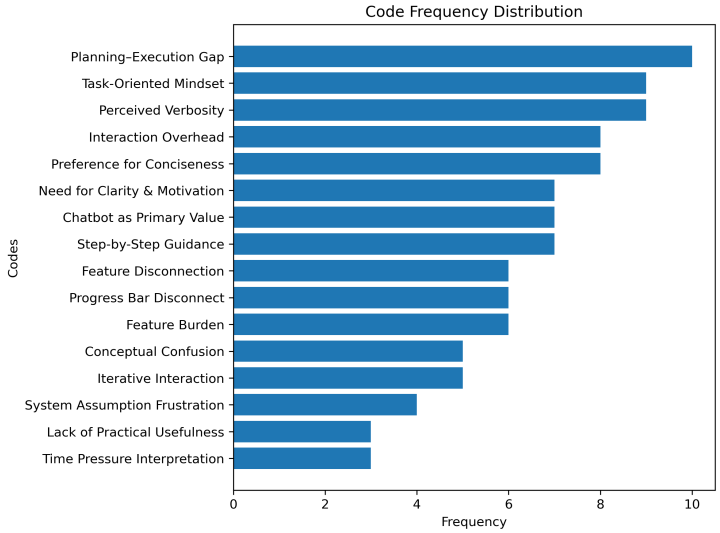

Figure 12: Frequency of coded instances across all participants, highlighting dominant issues such as actionability and verbosity.

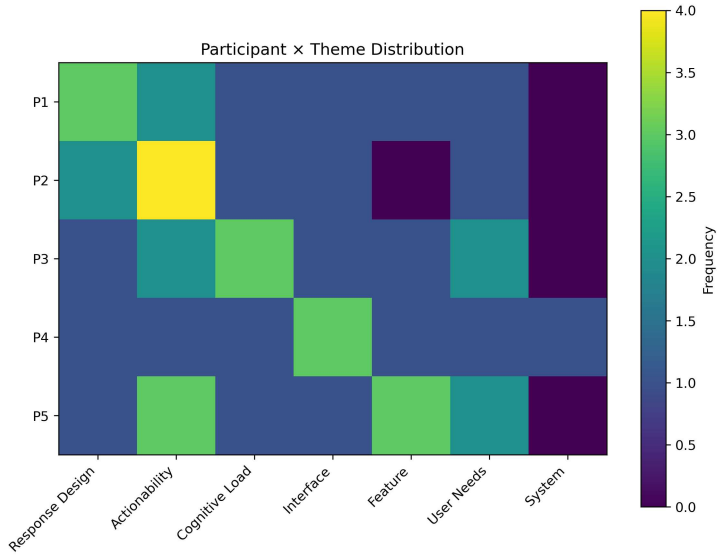

Figure 13: Participant-by-theme heatmap showing the distribution of interaction patterns.

Table 7: Theme-Level Aggregation

Theme	Codes	Total Instances
Actionability	C3–C5	22
Response Design	C1–C2	17
Cognitive Load	C6–C7	14
Interface Understanding	C8–C10	15
Feature Evaluation	C11–C13	16
User Needs	C14–C15	16
System Limitations	C16	3

1. Preference for Concise and Actionable Guidance (17 instances)
Participants consistently expressed a preference for short, direct, and actionable responses rather than long, reflective explanations. While the conversational interface aimed to follow motivational interviewing principles, several users found the responses overly verbose and cognitively demanding.
For example, one participant noted:
"My first thought is that it's too long... I just want to get it done."
Another participant described the responses as overwhelming:
"Seeing all that up front makes me not want to do it because it's a lot of stuff."
This suggests a mismatch between the system's reflective design and users' immediate need for task-oriented support, particularly in contexts where users are already experiencing avoidance.

2. Gap Between Reflection and Action Execution (22 instances)
A major breakdown identified was the system's inability to effectively transition users from reflection to concrete action. While participants appreciated the supportive tone, they struggled to derive clear, executable steps.
One participant explicitly highlighted this issue:
"The issue is not planning—the issue is execution."
Users preferred step-by-step, incremental guidance rather than full plans presented upfront. For instance:
"It would be better... just try and get me to move through one step at a time."
This indicates that effective support for procrastination requires not only reflection but also scaffolding action in small, manageable steps. This finding

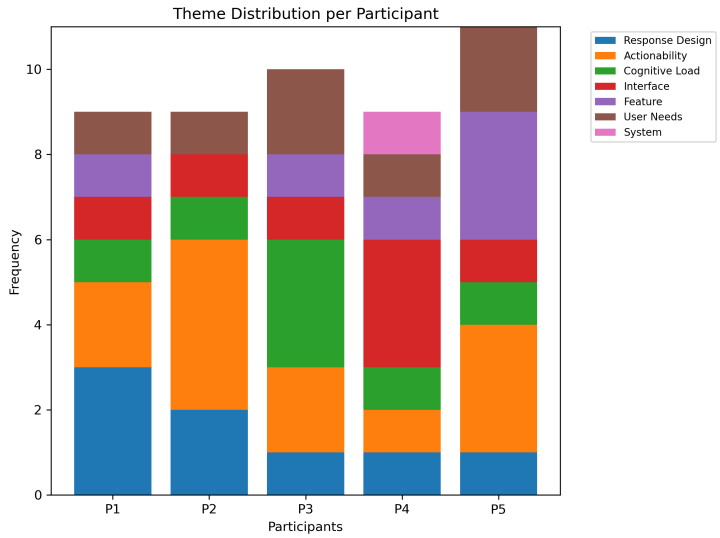

Figure 14: Stacked bar chart illustrating the composition of themes per participant, revealing variability in user behavior and interaction challenges.

suggests that the presence of action-planning tools alone may be insufficient; future systems should provide more proactive and timely support to help users convert intentions into concrete behavioral steps.

3. Interface and Conceptual Understanding Challenges (15 instances)

Several interface elements were not immediately intuitive to users, particularly the progress bar and abstract stages (e.g., "evoking").

One participant reported:

"I don't understand what evoking means... what does it mean to me?"

Similarly, the progress bar was perceived as disconnected from actual user progress from one participant:

"We're just having a conversation... the progress bar doesn't make sense to me."

These findings suggest that while theoretically grounded (e.g., motivational interviewing stages), such abstractions may lack transparency and perceived utility in practice.

4. Mixed Perception of Motivational and Gamification Features (14 instances)

Motivational features such as the reward tracker and tone of conversation interface received mixed feedback.
Positive perceptions:
Participants appreciated the supportive and empathetic tone of conversational interface:
"It seems understanding... it gave me some good advice."
Gamification elements (e.g., medals) were perceived as engaging:
"It's very helpful... especially for people who need motivation."
Negative perceptions:
However, these features were also seen as adding friction:
"The reward tracker requires me to type what I've done... it's more work."
Similarly, the Pomodoro timer was perceived as disconnected:
"I think it's quite disconnected... I still don't have clarity on what to do first."
This highlights a critical tension: motivational features must reduce, not increase, cognitive and interaction effort.

5. Conversational interface as the Primary Value Driver (7 instances)
Across all participants, the conversational interface interaction itself was consistently identified as the most useful component.
For example:
"The easiest for me to use was definitely the bot."
"The chat would be the most useful... as long as responses are shorter."
This suggests that conversational interaction remains the core mechanism of engagement, but its effectiveness depends heavily on response design.

6. Cognitive Load and Interaction Overhead (14 instances)
Participants frequently experienced cognitive overload when interacting with multiple features simultaneously (e.g., conversational interface, reward tracker, progress bar, timer). Rather than supporting task engagement, the accumulation of features sometimes increased friction.
For instance, users expressed reluctance to engage with features that required additional input or interpretation, especially when already feeling unmotivated. This suggests that even well-intentioned features can become burdensome if they demand extra effort.
Therefore, in low-motivation contexts such as procrastination, users are particularly sensitive to cognitive load. Systems should prioritize minimal inter-

action cost and avoid requiring users to process or input excessive information.

We also categorized coded instances into behavioral types:

Table 8: Behavior Type Distribution

Behavior Type	Instances	Proportion
Action-oriented	43	∼55%
Cognitive/effort-related	29	∼37%
Reflective/interpretive	6	∼8%

To better understand which components of the interface attracted participants' attention, we analyzed the frequency of references to interaction elements during the think-aloud sessions and interviews. The chatbot conversation was the most frequently discussed element, followed by the progress bar and planning tools. Participants often described the chatbot as the primary source of value, emphasizing its role in providing guidance, reflection, and motivation. The progress bar also attracted substantial discussion, particularly regarding how progress should be represented in open-ended conversational tasks. Planning-related features, such as task breakdown and implementation intentions, were frequently mentioned as useful mechanisms for translating intentions into concrete actions. In contrast, reward tracking, avatar customization, and the Pomodoro timer were discussed less often, suggesting that participants placed greater emphasis on features directly supporting task initiation and completion than on supplementary motivational elements.

Table 9 shows the frequency with which participants referenced different interaction elements during the evaluation. Conversational support, progress visualization, and planning tools dominated participant feedback, indicating that these components were central to the user experience.

6.8. Summary

The findings indicate that:

- **Actionability issues are the dominant challenge:** Future designs should leverage LLMs to automatically generate structured if–then

Table 9: Frequency of participant references to interaction elements.

Interaction Element	References
Chatbot Conversation	12
Progress Bar	10
Planning Tool	7
Reward Tracker	4
Avatar Customization	4
Pomodoro Timer	3
Progress Detection Messages	3

plans, reducing the effort required for users to translate intentions into concrete actions.

- **Verbosity and cognitive load are consistent barriers:** Responses should be concise, structured, and optimized for low-effort interaction.
- **Interface understanding remains a systemic issue:** Abstract concepts and system states should be made more transparent and aligned with users' mental models.
- **Chatbot interaction is effective but needs refinement:** The chatbot should prioritize actionable, step-by-step guidance to better support task initiation and execution.
- Conversational guidance, progress feedback, and planning support were central to participants' experiences with the system.

7. Discussion

This paper examined how conversational systems can support task engagement and reduce procrastination by combining reflective dialogue with action-oriented features. Across two studies, the findings reveal a consistent gap between reflection and execution, offering important insights for the design of LLM-powered behavior change systems.

7.1. From Reflection to Action

Consistent with prior work on Motivational Interviewing, participants engaged with the system's reflective prompts and reported increased awareness

of their procrastination behaviors. However, this awareness did not reliably translate into action. Instead, participants emphasized that the primary challenge lies in execution rather than planning. This finding highlights a limitation of reflection-focused conversational systems: while they can support understanding and motivation, they often fall short in facilitating task initiation. Our results suggest that effective systems must explicitly support the transition from reflection to action within the interaction itself.

7.2. Effectiveness of the Progress Bar

Findings from Study 1 indicate that the progress bar enhanced usability, confidence, perceived performance, and emotional responses, while slightly increasing effort and time pressure. The chatbot promoted reflection on procrastination but did not always lead to immediate action. While some participants perceived it as a useful structural cue that signaled conversation flow and proximity to completion, many either did not notice it or found it difficult to interpret in relation to their actual task progress. In several cases, participants noted that the progress indicator was most effective when linked to clearly defined goals, highlighting opportunities to further align progress feedback with users' evolving objectives in open-ended conversational settings. These results suggest that progress indicators, when applied to conversational interfaces, must clearly communicate what is being tracked and how it relates to user outcomes.

7.3. Managing Cognitive Load

Our findings also underscore the importance of minimizing cognitive and interactional overhead. Participants reported disengagement when faced with verbose responses, multiple simultaneous features, or unclear interface elements. This is particularly salient in the context of procrastination, where users may already experience reduced cognitive capacity for task initiation.

7.4. The Role of Conversational Interaction

Across both studies, the chatbot interaction itself was consistently identified as the most useful component. Participants valued the flexibility and immediacy of conversational engagement, reinforcing its role as a central interaction paradigm. At the same time, the effectiveness of this interaction depended strongly on response design. Responses that were concise, structured, and action-oriented were more likely to support engagement, whereas verbose or repetitive responses reduced usability.

7.5. Implications for Design

Taken together, our findings suggest several implications for the design of conversational behavior change systems. First, systems should tightly integrate reflection with actionable next steps, ensuring that each interaction supports task progression. Second, responses should be concise and structured to reduce cognitive load. Third, action guidance should be incremental and context-sensitive rather than comprehensive. Finally, interface elements should be transparent and directly relevant to user goals, avoiding abstract or ambiguous representations.

8. Conclusion

This paper investigated how conversational systems can support task engagement and reduce procrastination. From two studies, we found a persistent gap between reflection and action: while users valued reflective support, they primarily needed concise, actionable guidance to initiate tasks. Cognitive load and interaction overhead further limited engagement, highlighting the importance of low-effort design.

The progress bar showed limited effectiveness, as it was often unnoticed or poorly aligned with users' perceived progress. In contrast, conversational interaction emerged as the most valuable component, though its effectiveness depends on clear, brief, and action-oriented responses.

Overall, our findings suggest that conversational behavior change systems should prioritize actionability, minimize cognitive load, and provide transparent, meaningful feedback to better support the transition from intention to action.

9. Limitations and Future Work

This work has several limitations. First, the evaluation focused on short-term interactions, and therefore does not capture long-term effects on behavior change, habit formation or sustained engagement. Second, the system was evaluated in a controlled study setting, which may not fully reflect real-world usage where contextual factors and interruptions influence behavior. Additionally, the think-aloud protocol may have influenced user behavior, as participants verbalized their thoughts during interaction.

Future work should explore longitudinal deployments to examine how action-oriented conversational support influences sustained task engagement

and habit formation over time. Further research is also needed to develop adaptive systems that dynamically adjust response length, level of guidance, and feature presentation based on user state and context. In particular, integrating LLM-driven mechanisms to automatically generate structured, step-by-step action plans (e.g., if–then plans) may help reduce the gap between intention and execution. Finally, future designs should investigate more intuitive and transparent representations of system feedback, such as progress indicators that are tightly aligned with user goals and perceived progress.

10. Ethical Statement

The experimental protocol was reviewed and approved by the Research Ethics Committee at Loughborough University (approval number: 24304 and 25419) prior to data collection. Informed consent was obtained from all participants before their participation in the studies. Written consent was obtained from all participants.

11. Data Availability Statement

The data supporting the findings of this study are publicly available at https://zenodo.org/records/23697333

References

Ahmed, I., Rose, J., Keilty, E., Cooper, C., Selby, P., 2022. Generation and classification of motivational-interviewing-style reflections for smoking behaviour change using few-shot learning with transformers. techrxiv. Preprint posted online June 13, 10–36227.

Almalki, K., Alharbi, O., Al-Ahmadi, W., Aljohani, M., 2020. Anti-procrastination online tool for graduate students based on the pomodoro technique, in: International Conference on Human-Computer Interaction, Springer. pp. 133–144.

van Baal, S.T., Le, S.T.T., Fatehi, F., Verdejo-Garcia, A., Hohwy, J., 2022. Effecting behaviour change using an artificial intelligence chatbot: a pilot randomised controlled study .

Baumel, A., Fleming, T., Schueller, S.M., 2020. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. Journal of medical Internet research 22, e20631.

Bente, G., Rüggenberg, S., Krämer, N.C., Eschenburg, F., 2008. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human communication research 34, 287–318.

Bhattacharjee, A., Zeng, Y., Xu, S.Y., Kulzhabayeva, D., Ma, M., Kornfield, R., Ahmed, S.I., Mariakakis, A., Czerwinski, M.P., Kuzminykh, A., et al., 2024. Understanding the role of large language models in personalizing and scaffolding strategies to combat academic procrastination, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–18.

Chang, W.J., Chang, P.C., Chang, Y.H., 2024. The gamification and development of a chatbot to promote oral self-care by adopting behavior change wheel for taiwanese children. Digital Health 10, 20552076241256750.

Chen, B., Dethlefs, N., 2025. Proactimate: Evaluating llm-based chatbots for behavior change interventions .

Clifford, D., Curtis, L., 2016. Motivational interviewing in nutrition and fitness. Guilford Publications.

Deci, E.L., Ryan, R.M., 2000. The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry 11, 227–268.

Deterding, S., Dixon, D., Khaled, R., Nacke, L., 2011. From game design elements to gamefulness: defining" gamification", in: Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, pp. 9–15.

Fogg, B.J., 2002. Persuasive technology: using computers to change what we think and do. Ubiquity 2002, 2.

Fortunati, L., Manganelli, A.M., 2008. The social representation of telecommunications. Personal and Ubiquitous Computing 12, 421–431.

Gollwitzer, P.M., 1999. Implementation intentions: Strong effects of simple plans. American psychologist 54, 493.

Gollwitzer, P.M., Sheeran, P., 2006. Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in experimental social psychology 38, 69–119.

Harrison, C., Amento, B., Kuznetsov, S., Bell, R., 2007. Rethinking the progress bar, in: Proceedings of the 20th annual ACM symposium on User interface software and technology, pp. 115–118.

Hegde, N., Vardhan, M., Nathani, D., Rosenzweig, E., Speed, C., Karthikesalingam, A., Seneviratne, M., 2024. Infusing behavior science into large language models for activity coaching. PLOS Digital Health 3, e0000431.

Hettema, J., Steele, J., Miller, W.R., 2005. Motivational interviewing. Annu. Rev. Clin. Psychol. 1, 91–111.

Jääskeläinen, R., 2010. Think-aloud protocol. Handbook of translation studies 1, 371–374.

Jeong, D., Shin, S., Jeong, J., 2025. Conversation progress guide: Ui system for enhancing self-efficacy in conversational ai, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems.

Jörke, M., Sapkota, S., Warkenthien, L., Vainio, N., Schmiedmayer, P., Brunskill, E., Landay, J., 2024. Supporting physical activity behavior change with llm-based conversational agents. arXiv preprint arXiv:2405.06061 .

Kumar, T.A., 2024. Generation of Backward-Looking Complex Reflections for a Motivational Interviewing-Based Smoking Cessation Chatbot using GPT-4. Ph.D. thesis.

Locke, E.A., 1997. Self-efficacy: The exercise of control. Personnel psychology 50, 801.

Martinengo, L., Jabir, A.I., Goh, W.W.T., Lo, N.Y.W., Ho, M.H.R., Kowatsch, T., Atun, R., Michie, S., Tudor Car, L., 2022. Conversational agents in health care: scoping review of their behavior change techniques and underpinning theory. Journal of Medical Internet Research 24, e39243.

Meyer, S., Elsweiler, D., . Llm-based conversational agents for behaviour change support: A randomized controlled trial examining efficacy, safety, and the role of user behaviour. Safety, and the Role of User Behaviour .

Miller, W.R., Moyers, T.B., Ernst, D., Amrhein, P., 2003. Manual for the motivational interviewing skill code (misc). Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico .

Miller, W.R., Rollnick, S., 2012. Motivational interviewing: Helping people change. Guilford press.

Munnukka, J., Talvitie-Lamberg, K., Maity, D., 2022. Anthropomorphism and social presence in human–virtual service assistant interactions: The role of dialog length and attitudes. Computers in Human Behavior 135, 107343.

Nepal, S., Pillai, A., Campbell, W., Massachi, T., Heinz, M.V., Kunwar, A., Choi, E.S., Xu, X., Kuc, J., Huckins, J.F., et al., 2024. Mindscape study: Integrating llm and behavioral sensing for personalized ai-driven journaling experiences. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 8, 1–44.

Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., Peters, C., 2020. Engagement in human-agent interaction: An overview. Frontiers in Robotics and AI 7, 92.

Oinas-Kukkonen, H., 2010. Behavior change support systems: A research model and agenda, in: International Conference on persuasive technology, Springer. pp. 4–14.

Orpen, C., 1998. The causes and consequences of academic procrastination: A research note. Westminster Studies in Education 21, 73–75.

Rollnick, S., Miller, W.R., 1995. What is motivational interviewing? Behavioural and cognitive Psychotherapy 23, 325–334.

Schroeder, R., 2002. Social interaction in virtual environments: Key issues, common themes, and a framework for research, in: The social life of avatars: Presence and interaction in shared virtual environments. Springer, pp. 1–18.

Skinner, B.F., 1965. Science and human behavior. 92904, Simon and Schuster.

Sullivan, S., 2020. 16 habit and the human lifespan. Habits: Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory , 337.

Thakkar, N., 2009. Why procrastinate: an investigation of the root causes behind procrastination .

Zimmerman, B.J., 2000. Attaining self-regulation: A social cognitive perspective, in: Handbook of self-regulation. Elsevier, pp. 13–39.