Park *et al. Int J Educ Technol High Educ* (2025) 22:28
https://doi.org/10.1186/s41239-025-00527-z

International Journal of Educational Technology in Higher Education

RESEARCH ARTICLE **Open Access**

Enhancing academic stress assessment through self-disclosure chatbots: effects on engagement, accuracy, and self-reflection

Minyoung Park[1], Sidney Fels[2] and Kyoungwon Seo[1*]

*Correspondence: kwseo@seoultech.ac.kr

[1] Department of Applied Artificial Intelligence, Seoul National University of Science and Technology, Seoul, South Korea, 232 Gongneung-Ro, Gongneung-Dong, Nowon-Gu, 01811

[2] Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada

Abstract

Academic stress significantly affects students' well-being and academic performance, highlighting the need for more effective assessment methods to guide targeted interventions. This study investigates how self-disclosure chatbots—designed to share relevant experiences and thoughts—can enhance academic stress assessments by increasing student engagement, improving accuracy, and fostering deeper self-reflection. Two chatbot conditions were developed: a self-disclosure (SD) chatbot that used personal narratives to build empathy, and a non-self-disclosure (NSD) chatbot. In a randomized experiment with 50 university students, participants interacted with either the SD or NSD chatbot. Results showed that the SD chatbot elicited significantly higher engagement, as evidenced by longer session lengths (15.55 ± 5.92 min) and higher word counts (240 ± 114.02 words), compared to the NSD chatbot (11.31 ± 5.21 min; 162.38 ± 66.24 words). Assessment accuracy—evaluated by comparing results from the SISCO Inventory of Academic Stress with chatbot-generated evaluations—was slightly higher for the SD chatbot (0.936) than for the NSD chatbot (0.862), based on accuracy within a ± one-point deviation. Moreover, students who interacted with the SD chatbot reported deeper self-reflection and developed more actionable strategies for managing their stress. Overall, these findings illuminate the value of self-disclosure in chatbot-based assessments and highlight broader applications for addressing academic stress and mental health challenges in educational settings.

Keywords: Academic stress assessment, Chatbots, Self-disclosure, Engagement, Accuracy, Self-reflection

Introduction

Academic stress is a prevalent concern in higher education, posing significant risks to students' well-being and academic performance (Pascoe et al., 2020). Prolonged exposure to such stress can disrupt learning, reduce motivation, and hinder academic success (Alyahyan & Düştegör, 2020). Addressing this issue effectively requires a nuanced understanding of academic stress and its underlying causes (Steare et al., 2023). Comprehensive assessments that identify specific stressors, symptoms, and coping strategies offer clearer insights into students' stress levels, thereby guiding targeted interventions (Choi et al., 2024; Macías, 2006). Reflecting this need, interdisciplinary research in computer science, education, and related fields increasingly underscores the importance of robust stress assessments as a foundational step in promoting student well-being (Jin et al., 2023; Sung et al., 2023).

Traditional methods for assessing academic stress, such as self-report questionnaires, are popular for their simplicity and ease of administration (Cohen, 1997; Jiménez-Mijangos et al., 2023). One widely used tool is the SISCO Inventory of Academic Stress (SISCO-AS), developed by Macías (2007). However, these approaches often do not foster the depth of self-reflection necessary to capture complex or subtle stressors, leading to superficial or automatic responses (Ghosh et al., 2020). Without meaningful self-reflection, students may overlook the underlying factors of their stress—whether these stem from personal beliefs, academic pressures, or an accumulation of minor stressors (Hjeltnes et al., 2015). While professional counseling can facilitate advanced reflection through guided inquiry, it is often limited by high costs, time constraints, and lack of availability in many educational settings (Gladding, 2009; Hanley et al., 2017). These barriers highlight the urgent need for accessible, cost-effective, and scalable stress assessment methods that inspire thoughtful self-reflection.

Chatbot-based approaches have emerged as a promising alternative to traditional assessments, offering the potential to increase accessibility and student engagement through interactive, conversational interfaces (Haque & Rubya, 2023; Schillings et al., 2024; Seo et al., 2021a, b). By simulating human-like dialogue, chatbots can create a more comfortable environment for students to discuss stress-related issues (Belda-Medina & Kokošková, 2023; Park et al., 2024). Consequently, chatbots are increasingly integrated into stress assessment initiatives, functioning as low-cost, scalable supplements to conventional practices (Ulrich et al., 2024; van der Schyff et al., 2023). Nevertheless, many existing chatbot-based tools still mirror traditional survey limitations by relying on a predefined set of questions that fail to encourage deeper self-reflection (Jin et al., 2024; Toulme et al., 2023). Realizing the full potential of chatbots for stress assessment thus requires designing experiences that prompt more meaningful engagement and self-reflection (Booth et al., 2023; Seo et al., 2021a, b).

An emerging strategy involves incorporating chatbot self-disclosure, wherein the chatbot shares relevant information or thoughts to build empathy and rapport with users (Park et al., 2023). By offering relatable examples that resonate with common student challenges, self-disclosure chatbots can cultivate a sense of shared experience, prompting students to engage in deeper self-reflection (Lee et al., 2020). This approach aligns with social penetration theory, which posits that disclosures of personal or relatable experiences can strengthen connections and promote self-reflection (Altman, 1973). Through natural, conversational patterns, self-disclosure chatbots may help students articulate their experiences more effectively, potentially enhancing both the depth and accuracy of academic stress assessments (Degani et al., 2017; Pfeifer et al., 2009). While preliminary evidence suggests that chatbot self-disclosure can bolster engagement and promote deeper reflection, Schick et al. (2022) highlight the need for more empirical data on its capacity to improve stress assessment accuracy. To address this gap, the present study develops and evaluates chatbots with self-disclosure functionalities, examining their impact on student engagement, assessment accuracy, and self-reflection.

As chatbot-based assessments gain traction in academic and mental health contexts, concerns have emerged regarding biases and unintended consequences in their interactions. Prior studies have identified risks such as the reinforcement of stigma and the potential for chatbot responses to negatively impact users' well-being (Petzel & Sowerby, 2025). To mitigate these risks, recent research has emphasized the importance of structured conversation flows (Xue et al., 2023), iterative refinement through pilot studies (Arun et al., 2024), and real-time monitoring for detecting unintended biases (Idowu et al., 2024). Building on prior research, this study employs a structured, human-centered design—incorporating prompt engineering, iterative pilot testing, and real-time monitoring—to develop a self-disclosure chatbot that promotes meaningful self-reflection while ensuring safe, unbiased, and accurate academic stress assessments. While our approach relied on manual refinement and pilot testing, future research could enhance scalability and precision in bias mitigation by integrating artificial intelligence-based bias detection methods to automate monitoring and further refine chatbot responses (Dentella et al., 2024). Three research questions guide this inquiry:

- RQ1: Do self-disclosure chatbots lead to greater student engagement compared to non-self-disclosure chatbots?
- RQ2: Do self-disclosure chatbots result in more accurate academic stress assessments than non-self-disclosure chatbots?
- RQ3: How do self-disclosure chatbots differ from non-self-disclosure chatbots in fostering self-reflection during academic stress assessments?

The paper is organized as follows: Sect. 2 reviews the existing literature on chatbot-based stress assessments and the role of self-disclosure. Section 3 describes the research design, including participant recruitment, measurement instruments, the two chatbot conditions, and data analysis methods. Section 4 presents the results from both quantitative and qualitative perspectives. Finally, Sect. 5 discusses the findings in relation to existing literature, addresses limitations, and outlines directions for future research.

Related work

Chatbot-based stress assessment

With the rapid advancement of Large Language Models (LLMs) such as ChatGPT, chatbots have become increasingly popular for stress assessment, offering an alternative to traditional methods like self-report questionnaires and professional counseling (Cha et al., 2024; Habicht et al., 2024). By enabling conversational and intuitive interactions, LLMs facilitate broader accessibility and engagement, prompting researchers and practitioners to explore more effective chatbot designs for stress assessment (Ahmed et al., 2023; Seo et al., 2024; Song et al., 2024).

Recent research on chatbot-based stress assessments frequently highlights the goal of automating processes to alleviate financial and logistical challenges associated with established methods. Although self-report questionnaires are efficient, they often fail to capture the nuanced responses necessary for greater insight (Cohen, 1997; Ghosh et al., 2020), and counseling—while comprehensive—is time-intensive and costly (Gladding, 2009; Hanley et al., 2017). Consequently, many studies have examined chatbots as scalable solutions, administering traditional self-report questionnaires in a digital, user-friendly format. For instance, Jin et al. (2024) integrated the UCLA Loneliness Scale into a chatbot, allowing students to complete the scale via conversation. Although participants reported higher satisfaction with the conversational approach, the chatbot merely replicated standard questionnaire items, limiting the potential for deeper self-reflection. Similarly, Toulme et al. (2023) developed a chatbot that administered the Patient Health Questionnaire for depression screening without incorporating natural conversation. While this chatbot replicated the assessment successfully, it did not promote user comfort or self-reflection. These cases underscore the challenge of implementing chatbots that go beyond standardized questions to foster the meaningful reflection essential for accurate stress evaluations.

Some efforts have specifically addressed improving the quality of chatbot interactions. Kang and Kang (2024), for example, explored the role of visual cues and conversational styles in enhancing user engagement, finding that empathetic dialogue had a stronger impact than visual cues. Further, Lee et al. (2024) investigated the effect of empathetic expressions—such as sharing related personal thoughts—to increase social presence. Their findings suggested that chatbots offering relatable experiences delivered greater conversational depth and user comfort.

Building on these efforts, recent studies have identified self-disclosure in chatbot interactions as a powerful way to encourage self-reflection (Tsumura & Yamada, 2023; Zhang et al., 2024). Self-disclosure refers to the gradual sharing of relevant information or thoughts, a core mechanism for building trust and facilitating richer interactions (Ignatius & Kokkonen, 2007). Grounded in social penetration theory (Altman, 1973; Barak & Gluck-Ofri, 2007), this approach maintains that relationships strengthen as interactions progress from superficial to increasingly personal. In support of this idea, Lee et al. (2020) observed that chatbots sharing relatable challenges prompted reciprocal self-disclosure from users. Croes et al. (2023) similarly found that self-disclosure chatbots fostered greater trust, empathy, and conversational depth, while Park et al. (2023) demonstrated that self-disclosure increased user engagement and satisfaction.

As self-disclosure chatbots gain increasing attention, developing self-disclosing language models has become crucial for enhancing their effectiveness. Several models have been designed to incorporate structured self-disclosure, refining their ability to foster engagement and user trust. For example, Goddard et al. (2024) introduced Mibi, a self-disclosing chatbot designed to support computer science students through pre-scripted, empathetic responses, effectively reducing their fear of judgement. Kim (2024) examined the impact of disclosure type and chatbot identity on user trust, revealing that emotional self-disclosure in human-like chatbots significantly increased feelings of intimacy and connection. Liang et al. (2024) extended this research by categorizing self-disclosure into factual, cognitive, and emotional levels and implementing adaptive disclosure strategies, enhancing user engagement and reciprocal disclosure. However, these studies primarily focused on improving conversational flow and emotional connection rather than expanding self-disclosing chatbots to evaluative tasks.

More recently, research has further demonstrated the potential of self-disclosure chatbots in various applications. Cui et al. (2024) combined self-disclosure with non-stigmatizing interpretation to reduce mental illness stigma, identifying this combination as the most effective approach. Chung and Kang (2023) explored reciprocal self-disclosure in discussions of painful experiences, finding that it increased trust, enhanced conversational enjoyment, and provided users with a sense of relief post-interaction. Additionally, Jo et al. (2024) examined the long-term effects of self-disclosure in LLM-driven chatbots with memory systems for public health interventions, observing heightened user disclosures and sustained engagement.

Despite these advancements, relatively few studies have explored the application of self-disclosure chatbots in evaluative contexts, particularly in academic stress assessment. While existing research has largely focused on stigma reduction, emotional support, and user self-disclosure, questions remain regarding how self-disclosure can enhance assessment quality. To address this gap, our study applies self-disclosure within a structured conversation on academic stress, positioning self-disclosing chatbots not only as conversational agents but also as tools for stress assessment. By integrating the SISCO-AS into the chatbot conversation, we evaluate whether a self-disclosure chatbot can not only increase student engagement but also improve the accuracy of stress assessments and foster deeper self-reflection compared to a non-self-disclosure chatbot.

Self-reflection framework

Self-reflection, defined as the critical examination and evaluation of one's own thoughts, emotions, and behaviors, plays a pivotal role in accurately assessing stress (Grant et al., 2002). Recognizing its importance, researchers have developed a range of structured frameworks to guide self-reflection (Boud et al., 2013; Gibbs, 1988; Kolb, 2014). Among these, Koole et al. (2011) propose a widely used, comprehensive model that categorizes self-reflection into three stages—*reviewing the experience*, *critical analysis*, and *reflective outcome*—each comprising two substages (see Table 1). This structured, step-by-step approach has become an essential tool for studying self-reflection in academic stress assessment.

In the first stage of Koole et al.'s (2011) framework, *reviewing the experience*, individuals move from a factual description of an event—presenting it objectively and free from interpretation—to awareness, where they acknowledge the thoughts, emotions, and contextual factors that shaped the experience. These substages lay a strong foundation for advanced reflection by maintaining a clear, unbiased understanding of what occurred.

Table 1 Self-reflection framework, comprising three stages and six substages to guide the self-reflection process, adapted from Koole et al. (2011)

Stage	Substage	Description
Reviewing the experience	Description	Describing the event or situation objectively without interpretation or judgment
	Awareness	Identifying key elements of the experience, including thoughts, emotions, and contextual factors
Critical analysis	Reflective inquiry	Asking probing questions to explore the event in depth
	Processing in a frame of reference	Interpreting the event through one's beliefs, values, and assumptions
Reflective outcome	New perspectives	Gaining new insights or viewpoints based on the reflective process
	Reflective behavior	Translating insights into actionable strategies for future situations

The second stage, *critical analysis*, promotes a more thorough understanding of the event and one's reactions. This stage begins with reflective inquiry, in which individuals pose questions to identify causes, implications, and personal responses to the event. It then progresses to processing in a frame of reference, where the event is interpreted through personal beliefs and assumptions, enabling a more comprehensive view and insight.

The final stage, *reflective outcome*, emphasizes applying insights to real-life situations. During new perspectives, individuals reformulate their understanding of the event, potentially altering how they respond to similar situations in the future. This leads to reflective behavior, which focuses on implementing these insights as actionable strategies for change, ensuring that lessons learned translate into tangible improvements in decision-making and stress management.

In the context of academic stress assessment, the depth of self-reflection significantly influences both student engagement and the accuracy of evaluations. Students who engage only in superficial reflection (e.g., merely recounting events) may provide less accurate responses regarding their stress levels. Conversely, those who reach deeper levels of reflection—including critical analysis and reflective outcomes—are more likely to offer detailed, thoughtful assessments (Alaoutinen, 2012). Consequently, encouraging more advanced stages of self-reflection is essential for obtaining meaningful evaluations of academic stress. Chatbots designed to promote this form of self-reflection could help students assess their stress more accurately (Park & Seo, 2023; Travers et al., 2015; Xiao et al., 2020).

Building on Koole et al.'s work (2011), Ahlqvist-Björkroth et al. (2023) introduced a qualitative approach for assessing individuals' levels of self-reflection by analyzing interview data. Adapting this method, the present study conducted post-interaction interviews with students who engaged with either a self-disclosure or non-self-disclosure chatbot. By applying Koole's framework to these interviews, we aimed to examine how effectively SD chatbots foster deeper self-reflection in academic stress assessments compared to NSD chatbots.

Materials and methods

Participants

Fifty university students with at least one year of higher education experience were recruited from a human–computer interaction course based on availability and willingness to participate. The sample size was determined via a power analysis, aimed at detecting a medium effect size with sufficient statistical power ($\alpha = 0.05$). This analysis indicated a minimum requirement of 44 participants; to account for potential dropouts and ensure adequate power, 50 participants were recruited. No participants dropped out during the experiment. Informed consent was obtained prior to the study.

Participants were randomly assigned to one of two conditions: the non-self-disclosure (NSD) chatbot ($n = 25$) or the self-disclosure (SD) chatbot ($n = 25$). In the NSD condition, 10 participants were male and 15 were female (mean age = 22.6, SD = 1.68). In the SD condition, 12 participants were male and 13 were female (mean age = 22.28, SD = 1.43). There were no statistically significant differences between conditions in terms of gender distribution ($\chi^2 = 0.08$, $p = 0.78$) or age ($t = 0.72$, $p = 0.47$). All study procedures, including participant recruitment, were approved by the Institutional Review Board at Seoul National University of Science and Technology (IRB approval number: 2023-0021-01).

Academic stress assessment questionnaire

Academic stress was evaluated using the SISCO Inventory of Academic Stress (SISCO-AS; Macías, 2006, 2007), a widely employed self-report questionnaire that measures three dimensions of academic stress: stressors, symptoms, and coping strategies. The full instrument contains 31 items, with the first item confirming eligibility (yes/no) and the second assessing overall academic stress intensity. The remaining 29 items target the primary dimensions of stress: eight focus on stressors, 15 on stress-related symptoms, and six on coping strategies. All items use a 5-point Likert scale ranging from 1 (never) to 5 (always). For this study, only the 29 items measuring stressors, symptoms, and coping strategies were analyzed; the first two items (eligibility and general stress intensity) were excluded. These 29 items can be accessed at https://osf.io/ch96q/?view_only=2f02931fbfe7482bae2b21fcbb5e9213.

Self-report questionnaire as ground truth

To establish a reliable ground truth for academic stress assessment, we used participants' SISCO-AS scores as the baseline measure. However, recognizing the inherent limitations of self-reported data, we implemented two key enhancements to improve reliability and minimize variability. First, to mitigate inattentive or rushed responses, participants were explicitly instructed to carefully read and consider each question, ensuring thoughtful and deliberate answers—an approach aligned with best practices for enhancing self-report accuracy (Ward & Meade, 2023). Second, to enhance clarity and reduce ambiguity, each of the 29 items was supplemented with a brief example sentence. For instance, when assessing the frequency of intense competition, we provided an example such as "Feeling stressed due to fierce competition with peers over grades," helping participants interpret the question more precisely. This modification follows established guidelines for improving clarity in self-report questionnaires (Neijens et al., 2024). These measures aimed to enhance the validity and reliability of the self-reported responses, thereby strengthening the accuracy of the ground truth used in this study.

Two different chatbot conditions

To examine the impact of self-disclosure in chatbots on academic stress assessment, we developed two chatbot conditions: a NSD chatbot and a SD chatbot. Both chatbots were created using prompt engineering and implemented through Streamlit, utilizing the GPT-4o API. Their responses were carefully designed to maintain a neutral, supportive tone by adopting a friendly, empathetic, peer-like persona while focusing exclusively on academic stress. To ensure unbiased interactions, all chatbot responses were structured to encourage open-ended self-reflection without imposing opinions on participants' experiences, thereby avoiding generalized or stigmatizing language (Lee et al., 2023). Each chatbot used questions from the SISCO-AS to assess academic stress, with variations in their interaction style. The NSD chatbot maintained a task-focused tone, strictly evaluating participants' stress levels without sharing any personal information, similar to traditional chatbot-based stress assessments. In contrast, the SD chatbot incorporated self-disclosure techniques, sharing relatable information or thoughts before posing key questions to foster rapport and encourage participants' self-reflection. Participants accessed the chatbots via links on their personal laptops in a private, isolated experiment room, ensuring both comfort and confidentiality.

To clearly distinguish self-disclosure from task-focused interactions and minimize subjective influences, we implemented a structured conversation flow (Xue et al., 2023) consisting of four distinct phases: (1) small talk, (2) self-disclosure (SD chatbot only), (3) key questions, and (4) closing. During (1) small talk, both chatbots greeted participants with, "Hi! How's your day going so far?" After receiving participant responses, only the SD chatbot moved to the (2) self-disclosure phase, sharing experiences such as, "I've been struggling to remember information even though I use flashcards, and it's been frustrating," to mirror potential participant experiences and promote self-reflection. The NSD chatbot skipped this phase, proceeding directly to (3) key questions, where both chatbots assessed academic stress using SISCO-AS-based questions (see Fig. 1). In this phase, the SD chatbot incorporated self-disclosure into its questions, asking participants, "I often feel overwhelmed during exams. Do you experience this too?" Conversely, the NSD chatbot posed straightforward questions about academic stress without self-disclosure. Finally, in the (4) closing phase, both chatbots thanked the participants for their time. Before deployment, a pilot study was conducted to refine the chatbot prompts and mitigate unintended biases. Multiple simulations and participant feedback ensured a consistently neutral tone, and anonymized interaction logs throughout the study confirmed that no signs of bias were present.

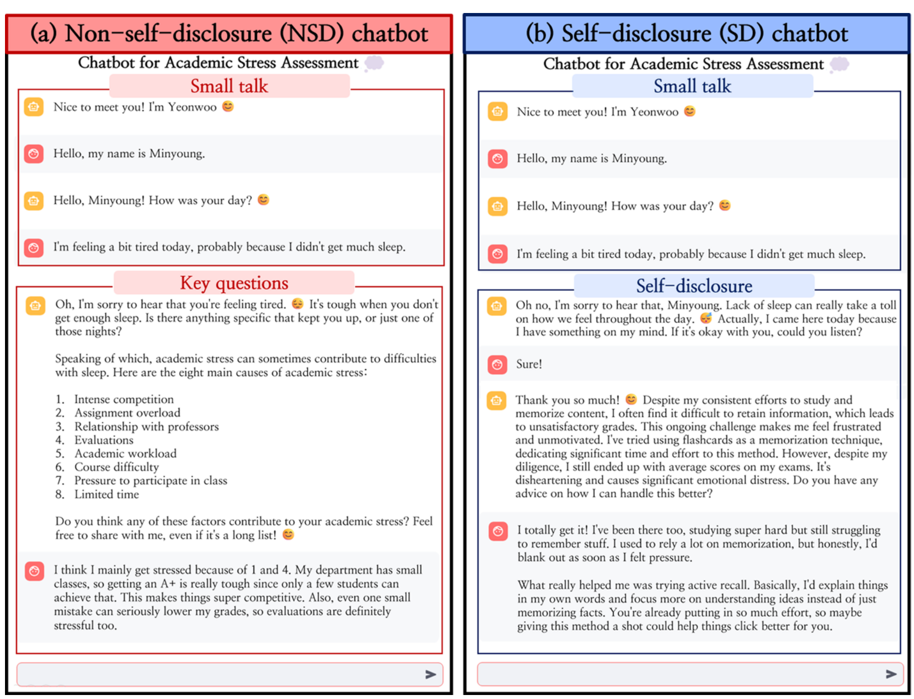

Fig. 1 Screenshots illustrating the conversation phases in the chatbot interactions: **a** The non-self-disclosure (NSD) chatbot transitions directly from small talk to key questions; **b** The self-disclosure (SD) chatbot includes a self-disclosure phase between small talk and key questions

Upon completion of the chatbot interaction, log data were analyzed to measure students' engagement and academic stress. Engagement was assessed using metrics from the chatbot logs, including session length and word count. Session length, recorded in seconds, represented the time participants spent interacting with the chatbots, with longer session length indicating higher engagement (Mead, 2019). Similarly, word count served as an indicator of interaction depth, with a higher word count suggesting greater participant involvement in discussing their academic stress (Van Doorn et al., 2010). These metrics allowed for a comprehensive analysis between the NSD and SD chatbot conditions to assess the impact of self-disclosure on students' engagement.

For academic stress assessment, participants' conversational input with the chatbots was analyzed using the GPT-4o API. Through prompt engineering, the LLM scored responses based on the 29 items in the SISCO-AS, using a 5-point Likert scale from 1 (never) to 5 (always). The use of LLMs to evaluate conversation logs is a widely accepted method in recent studies (Lundgren, 2024). By comparing the academic stress scores generated by the LLM with participants' original SISCO-AS responses, we evaluated the accuracy of stress assessment in each chatbot condition. Detailed prompt engineering instructions for both the NSD and SD chatbots, including conversation protocols and stress evaluation methods, can be accessed at https://osf.io/ch96q/?view_only=2f02931fbfe7482bae2b21fcbb5e9213.

Experimental procedures

Participants completed the SISCO-AS survey and interacted with the chatbot in a counterbalanced order to prevent order effects, ensuring that the sequence of activities did not impact the outcomes. Before beginning the chatbot conversation, participants were informed that the interaction would involve a stress assessment. The experiment was conducted in a quiet, private setting to allow participants to focus fully on the tasks. Although participants were informed that they could terminate the experiment at any time if they felt uncomfortable, none chose to do so.

Following the SISCO-AS survey and chatbot interaction, participants took part in a semi-structured interview to explore the self-reflection processes that the chatbot experience may have initiated. The interview aimed to capture detailed insights into how the chatbot supported students' self-reflection on academic stress and whether their perspectives changed as a result of the interaction. Example questions included: "Can you describe your overall experience with the chatbot?", "How did the chatbot help you reflect on your academic stress?", "Did interacting with the chatbot change your thoughts about academic stress?", "Were there any particular chatbot messages that you found especially thought-provoking or helpful?", and "Did sharing or hearing personal experiences during the session affect your level of self-reflection?" The entire experimental session lasted approximately 46.85 ± 11.19 min, with participants spending an average of 13.43 ± 5.92 min interacting with the chatbot, around 10 min on the SISCO-AS survey, and about 25 min in the interview.

Data analysis

To comprehensively evaluate the study, we analyzed three key elements: student engagement, stress assessment accuracy, and self-reflection, each measured through specific methods and metrics. First, student engagement was assessed by examining session length and word count during chatbot interactions. Independent samples t-tests were conducted to compare session length (measured in seconds) and word count between the NSD and SD chatbot conditions, determining whether these engagement metrics differed significantly between the two groups.

Second, stress assessment accuracy was evaluated by comparing the 5-point scores predicted by the GPT-4o model (based on participants' conversation logs) with participants' SISCO-AS results. The chatbot's assessment performance was measured using precision (the proportion of correctly identified stress cases among all identified as stressed), recall (the chatbot's ability to identifying cases of academic stress), F1 score (the harmonic mean of precision and recall, balancing both metrics), accuracy (the overall percentage of correct matches between chatbot predictions and SISCO-AS results), and ACC1 (a metric that allows a one-point deviation between predicted and actual scores to still be considered accurate; Gaudette & Japkowicz, 2009). Macro-averaging was applied for precision, recall, and F1 score to ensure equal consideration of all stress levels, regardless of class size, providing a balanced evaluation (Sokolova & Lapalme, 2009). The inclusion of ACC1 allowed for a nuanced assessment by acknowledging minor discrepancies while maintaining emphasis on accuracy. Differences in these metrics between the NSD and SD chatbot conditions were analyzed using Barnard's test.

Lastly, self-reflection was assessed through narrative analysis of participants' responses in the semi-structured interviews (Bamberg, 2012). While generative AI tools can assist with preliminary text analysis, they currently lack the ability to fully interpret nuanced contextual and emotional cues essential for rigorous narrative analysis (Dentella et al., 2024; Farber, 2025). Additionally, ethical considerations necessitate human oversight to ensure participant confidentiality and maintain adherence to ethical research standards (Bouhouita-Guermech et al., 2023). Given these factors, human researchers conducted the narrative analysis to preserve the depth, integrity, and ethical handling of participants' reflections. This approach captured how students' self-reflections progressed through the stages and substages of Koole et al.'s framework (2011; see Sect. 2.2), allowing a qualitative comparison of self-reflection depth. Initially, the first author coded the narratives according to the stages and substages, capturing the flow and depth of each participant's story. Then, two other authors independently reviewed and refined this coding. Through four rounds of discussion, the final narratives were confirmed by unanimous agreement among the three authors, ensuring they accurately represented students' self-reflection processes and aligned with the study's research questions.

Results

Student engagement

Student engagement with the chatbot, as measured by session length and word count, showed statistically significant differences between the NSD and SD chatbot conditions (see Table 2). Participants in the SD chatbot condition had a longer average session length of 15.55 ± 5.92 min, compared to 11.31 ± 5.21 min in the NSD chatbot condition ($p < 0.05$), suggesting that the self-disclosure features in the SD chatbot encouraged participants to spend more time interacting. Additionally, participants in the SD chatbot condition produced a significantly higher word count (240 ± 114.02) than those

Table 2 Comparison of student engagement between non-self-disclosure (NSD) and self-disclosure (SD) chatbot conditions

Student engagement	NSD chatbot condition	SD chatbot condition	*p* value
Session length (minutes)	11.31 ± 5.21	15.55 ± 5.92	$p < 0.05$
Word count (words)	162.38 ± 66.24	240 ± 114.02	$p < 0.01$

Table 3 Comparison of assessment accuracy between non-self-disclosure (NSD) and self-disclosure (SD) chatbot conditions

Assessment accuracy	NSD chatbot condition	SD chatbot condition	*p* value
Precision	0.398	0.4	0.488
Recall	0.294	0.379	0.186
F1 score	0.302	0.377	0.191
Accuracy	0.398	0.516	0.053
ACC1	0.862	0.936	0.089

ACC1 accuracy within ± one-point deviation

in the NSD chatbot condition (162.38 ± 66.24; $p < 0.01$), indicating that they were more expressive and provided more detailed responses. These findings demonstrate that incorporating self-disclosure into the SD chatbot design significantly increased student engagement—evidenced by longer session durations and higher word counts—thereby underscoring its role in promoting extended conversations.

Stress assessment accuracy

Assessment accuracy metrics, derived by comparing participants' SISCO-AS results with the 5-point scores predicted by the GPT-4o model from conversation logs, showed that the SD chatbot slightly outperformed the NSD chatbot across all measures, though differences were not statistically significant (see Table 3). Precision was 0.398 for the NSD chatbot and 0.4 for the SD chatbot, while recall improved from 0.294 (NSD) to 0.379 (SD). The F1 score, a harmonic mean of precision and recall, increased from 0.302 (NSD) to 0.377 (SD). Accuracy, reflecting the percentage of exact matches between chatbot-derived stress scores and SISCO-AS results, rose from 0.398 (NSD) to 0.516 (SD). The ACC1 metric, allowing a margin of ± one-point, further highlighted the SD chatbot's advantage, improving from 0.862 (NSD) to 0.936 (SD). These results suggest that incorporating self-disclosure into chatbots can enhance the accuracy of stress evaluations, even if the improvement is modest.

Self-reflection

Table 4 summarizes our narrative analysis of participants' self-reflection experiences in both the NSD and SD chatbot conditions, with positive narratives denoted by (+) and negative narratives by (−). Across both conditions, participants generally reported positive experiences during the 'description' substage of the *reviewing the experience* phase, expressing comfort in sharing their thoughts with the chatbot. However, as the self-reflection process progressed—from the 'awareness' substage to the 'reflective behavior' substage—participants in the NSD condition reported more negative narratives,

Table 4 Summary of self-reflection on academic stress in non-self-disclosure (NSD) and self-disclosure (SD) chatbot conditions

Stage	Substage	NSD chatbot condition	SD chatbot condition
Reviewing the experience	Description	(+) The chatbot's neutral tone made participants feel safe, encouraging honest and open descriptions of their experiences	(+) The chatbot's supportive tone made participants feel understood, encouraging them to provide detailed descriptions of their experiences, thoughts, and feelings
	Awareness	(–) The chatbot's superficial empathy and one-sided questions limited participants' awareness, preventing deep reflection on their thoughts and feelings	(+) The chatbot's deep empathy and self-disclosure heightened participants' awareness of their thoughts and feelings, enabling deeper reflection
Critical analysis	Reflective inquiry	(–) The chatbot's rigid, formulaic questions led to quick, surface-level responses, limiting participants' ability to deeply analyze their stress	(+) The chatbot's empathetic and relatable questions encouraged participants to engage in reflective inquiry, leading to more meaningful contemplation of their academic stress
	Processing in a frame of reference	(–) The rigid, mechanical nature of the chatbot's questions made it difficult for participants to connect their experiences to their personal beliefs and values, resulting in superficial reflection	(+) The chatbot's self-disclosure helped participants connect their reflections to their personal beliefs and values, guiding them toward more meaningful reflection on academic stress
Reflective outcome	New perspectives	(–) The repetitive and shallow questions limited participants' ability to explore new insights, keeping them focused on familiar thoughts without developing new perspectives	(+) The chatbot's detailed and engaging questions helped participants uncover new insights, allowing them to gain fresh perspectives on their academic stress
	Reflective behavior	(–) The structured, formulaic conversation limited participants to reinforcing existing knowledge, preventing the development of new coping strategies or actionable plans	(+) The chatbot's reflective guidance encouraged participants to explore new coping strategies and apply insights to better manage their academic stress

suggesting that this chatbot was less effective in supporting advanced self-reflection. In contrast, participants in the SD condition consistently provided positive narratives across all stages of self-reflection, indicating that the SD chatbot's self-disclosure features facilitated a deeper and more meaningful reflective process. These findings suggest that self-disclosure in the SD chatbot enhanced participants' ability to engage fully with the self-reflection framework. A detailed example of our narrative analysis is available at https://osf.io/ch96q/?view_only=2f02931fbfe7482bae2b21fcbb5e9213.

Reviewing the experience

The *reviewing the experience* stage involves examining past events to identify and understand key aspects (Korthagen & Vasalos, 2005). This stage is divided into two substages: 'description' and 'awareness.' The 'description' substage involves recounting the experience objectively, laying a factual foundation for advanced reflection.

The ‘awareness’ substage focuses on recognizing and reflecting on the emotions and thoughts tied to the event, facilitating a greater understanding of its impact.

Description Participants in both the NSD and SD chatbot conditions reported positive narratives during the ‘description’ substage, though their reasons for feeling comfort differed. In the NSD condition, participants appreciated the chatbot’s neutral and non-judgmental tone, which encouraged open descriptions without fear of evaluation. One participant noted, for instance, that the chatbot “wouldn’t judge [her] regardless of what [she] said” (NSD #12), while another felt comfortable sharing personal information because the chatbot “wouldn’t take [her] concerns seriously” (NSD #1). This sense of neutrality fostered honest self-description.

Conversely, participants in the SD condition described the chatbot as friendly and supportive, creating an environment similar to confiding in a close friend. One participant explained, “it felt like chatting with a close friend, which made discussing [her] academic stress more natural” (SD #2). Another participant noted that the chatbot’s understanding eliminated any hesitation to open up (SD #16). Thus, while both chatbot conditions provided a comfortable environment, the NSD chatbot did so through neutrality, whereas the SD chatbot achieved it through warmth and support.

Awareness Clear differences emerged in the ‘awareness’ substage between the two chatbot conditions. Participants in the NSD condition generally did not reflect deeply on their thoughts and feelings, as the chatbot’s surface-level empathy and repetitive questions limited exploration. One participant described the chatbot’s interaction as “superficial and repetitive,” preventing her from addressing the personal aspects of her stress (NSD #4). Another participant remarked that “the chatbot’s responses felt artificial and didn’t allow for deeper thinking” (NSD #24), resulting in limited awareness and surface-level reflection.

In contrast, participants in the SD condition found that the chatbot’s empathy and self-disclosure facilitated heightened self-awareness. The chatbot’s empathetic questions and sharing prompted participants to explore their own thoughts and emotions more thoroughly. One participant noted, the chatbot “helped [her] recognize stressors [she] hadn’t considered before” (SD #2), while another said it “prompted [him] to think deeply and recall detailed experiences” (SD #21). Overall, the SD chatbot fostered greater awareness of academic stress, helping participants uncover insights that had previously gone unexamined.

Critical analysis

The *critical analysis* stage encourages deeper examination of experiences, prompting participants to question assumptions and contextualize their reflections within cognitive and emotional frameworks (Bourner, 2003; Mamede & Schmidt, 2004). This stage comprises two substages: ‘reflective inquiry,’ which focuses on formulating probing questions about the experience, and ‘processing in a frame of reference,’ where participants align reflections with personal beliefs and values.

Reflective inquiry Participants in the NSD and SD conditions showed distinct differences in how they approached 'reflective inquiry.' In the NSD condition, participants often provided quick, surface-level responses, with some acknowledging that while the chatbot made them think about their academic stress, it didn't foster greater understanding. As one participant explained, "the chatbot made me think about stress, but it didn't lead to deeper insights" (NSD #5). Another participant noted, "I responded quickly, like filling out a survey" (NSD #7), reflecting the NSD chatbot's structured, repetitive nature that limited in-depth inquiry.

In contrast, the SD chatbot encouraged participants to engage in a more meaningful reflective process. The chatbot's self-disclosure and relatable questions prompted participants to ask themselves more profound questions like "Have I experienced something similar?" or "How did I feel in that situation?". One participant shared, "hearing the chatbot's related information made me think, 'I've been through that too,' which helped me reflect more deeply" (SD #7). By relating the chatbot's examples to their own experiences, participants in the SD condition were able to delve into a more comprehensive exploration of their academic challenges (SD #24). Overall, the SD chatbot promoted more introspective questioning.

Processing in a frame of reference Notable differences also emerged between the two conditions in how participants contextualized their experiences within their personal beliefs and values. Participants in the NSD condition struggled to connect their reflections to their own frameworks, as the chatbot's repetitive questions felt impersonal and disconnected. One participant remarked, "the chatbot's questions felt invasive and didn't help me think about my stress in a personal way" (NSD #4), while another participant expressed frustration that, "the chatbot kept asking questions without sharing anything in return, making the conversation feel one-sided" (NSD #19). This limited depth in reflection prevented participants from aligning their experiences with personal values.

Conversely, participants in the SD condition found the chatbot's self-disclosure and relatable stories useful for guiding their reflections within a personal context. One participant explained, the chatbot's examples "helped me think about how to approach my own situation" (SD #5). This reciprocal interaction allowed participants to relate the chatbot's experiences to their own lives, resulting in more nuanced interpretation. Another participant noted, "the chatbot's stories made me recall my own experiences and helped me reflect on them in a more meaningful way" (SD #8). Overall, the SD chatbot facilitated a stronger connection between reflections and personal values, promoting a more contextualized understanding of academic stress.

Reflective outcome

The *reflective outcome* stage synthesizes insights gained from reflection, helping individuals translate these insights into actionable plans for future behavior and decision-making (Atkins & Murphy, 1993; Stockhausen, 1994). This stage includes two substages: 'new perspectives,' where individuals recognize previously overlooked insights, and 'reflective behavior,' where new understandings are applied to guide future actions.

New perspectives Participants in the NSD and SD conditions differed in their ability to gain new insights about academic stress. Many participants in the NSD condition felt that their interaction with the chatbot reinforced existing thoughts rather than leading to fresh perspectives. One participant remarked, “it helped [him] recognize existing thoughts, but it didn’t lead to new perspectives” (NSD #5), while another stated, it “didn’t provide any deep understanding of [her] stress situation” (NSD #4). This suggests that the NSD chatbot’s structured approach did not encourage exploration of new dimensions of stress.

In contrast, participants in the SD condition reported gaining clearer and more detailed insights into their academic stress. The chatbot’s relatable examples and engaging questions helped participants uncover new perspectives on stress causes and coping strategies. One participant shared, “the chatbot provided examples of stress types and coping mechanisms, helping [her] think more deeply about [her] own stress” (SD #4). Another participant said, “the conversation with the chatbot helped me realize aspects of stress that I hadn’t considered before” (SD #7). These findings indicate that the SD chatbot enabled a more thorough exploration, allowing participants to develop insights they had previously overlooked.

Reflective behavior Differences also emerged in the ability to translate reflections into actionable plans. Participants in the NSD condition found the chatbot useful for stress description but lacking in practical application. As one participant explained, “the chatbot helped me to describe my stress, but it didn’t help me develop new strategies for managing stress” (NSD #7). Another participant noted that while the chatbot helped him describe stress levels, “it didn’t offer insights into coping strategies” (NSD #13). The structured, superficial nature of the NSD chatbot’s interaction limited the development of actionable plans.

Conversely, participants in the SD condition reported success in translating reflections into practical strategies for managing stress. The chatbot’s reflective guidance helped participants identify both existing and new coping methods. One participant noted, “discussing specific stress situations helped [her] understand [her] coping methods and think about applying them more effectively” (SD #4). Another participant remarked that the chatbot’s questions expanded his view of coping options, saying, “I realized I have more ways to relieve stress than I thought” (SD #7). Participants in the SD condition consistently reported that the chatbot encouraged more thorough thinking about stress management, prompting them to consider a broader range of coping strategies.

In summary, participants’ interactions with the NSD and SD chatbots revealed significant differences across the substages of self-reflection. In the ‘description’ substage, both chatbot conditions made participants feel comfortable sharing, but the SD chatbot’s empathetic approach encouraged more detailed descriptions. During the ‘awareness’ substage, NSD participants found the interaction surface-level, while SD participants, prompted by relatable stories, engaged in heightened self-awareness. In the ‘reflective inquiry’ substage, NSD participants provided brief responses, while SD participants explored their academic stress more thoroughly. In the ‘processing in a frame of reference’ substage, NSD participants struggled to contextualize reflections, while SD participants effectively connected experiences to personal beliefs through the chatbot’s guidance. Regarding ‘new perspectives,’ NSD participants largely organized existing thoughts without gaining fresh insights, whereas SD participants developed new understandings of their stress. Finally, in the 'reflective behavior' substage, NSD participants reinforced existing knowledge but did not develop actionable plans, while SD participants identified and explored coping strategies for future application. These findings highlight the SD chatbot's superior effectiveness in promoting deeper self-reflection and actionable insights.

Discussion and conclusion

The primary aim of this study was to examine the impact of self-disclosure in chatbots on student engagement, accuracy of academic stress assessments, and depth of self-reflection during stress assessments. Two chatbot conditions were developed: a NSD chatbot and a SD chatbot. Findings indicate that the SD chatbot significantly enhanced student engagement, marginally improved stress assessment accuracy, and fostered a deeper level of self-reflection compared to the NSD chatbot.

In terms of student engagement, participants interacting with the SD chatbot showed significantly longer session length and higher word count than those using the NSD chatbot. This heightened engagement is likely due to the self-disclosure elements, which made the interaction feel more relatable and engaging. This finding aligns with prior research suggesting self-disclosure enriches conversation and fosters enhanced engagement (Lee et al., 2020). Our study extends this understanding by demonstrating the specific value of self-disclosure in academic stress assessments, as reflected in increased student engagement.

Although the differences in stress assessment accuracy between the SD and NSD chatbots were not statistically significant, the SD chatbot demonstrated a slight advantage in accurately predicting students' stress scores. This marginal improvement may be attributed to the more detailed and honest responses encouraged by the SD chatbot's empathetic approach. The high ACC1 value (0.936), which accounts for minor deviations, indicates that the SD chatbot aligns well with students' self-reported stress levels. These findings suggest that the SD chatbot's ability to promote richer reflections can enhance the quality of stress assessments, even if the improvement remains modest.

The most notable impact of self-disclosure was observed in self-reflection. Narrative analysis revealed that the SD chatbot significantly enhanced students' ability to engage in deeper self-reflection. By sharing its own experiences, the SD chatbot fostered an empathetic and supportive environment, prompting students to explore their thoughts and emotions more thoroughly. This approach led to more meaningful reflections on academic stress, enabling students to gain new perspectives and develop actionable strategies for stress management. In contrast, interactions with the NSD chatbot remained at a surface level, primarily within the 'description' substage, limiting students' progression through the full stages of self-reflection. This progression highlights that the SD chatbot, through self-disclosure, effectively guided students through each stage of self-reflection, while the NSD chatbot left students confined to the initial descriptive level.

Overall, these findings highlight the crucial role of self-disclosure in enhancing student engagement, assessment accuracy, and self-reflection within chatbot-based academic stress assessments. The results suggest that self-reflection may be central to improving both engagement and assessment accuracy. The SD chatbot's ability to encourage advanced self-reflection likely contributed to longer session length, higher word count, and marginal improvements in stress assessment accuracy. As LLMs depend on user interaction data to predict stress scores, richer conversational data can lead to greater assessment precision. These insights underscore the practical and theoretical importance of self-disclosure in designing effective chatbot-based academic stress assessments.

Theoretical implication

This study provides valuable theoretical insights into the role of self-disclosure in chatbot interactions, particularly within the field of educational technology in higher education. The findings suggest that self-disclosure chatbots can significantly enhance user engagement and facilitate deeper self-reflection, supporting theories in human–computer interaction that emphasize the impact of personal and relatable communication in digital platforms. Specifically, this study demonstrates that chatbots, by sharing relatable experiences, can create an environment conducive to self-reflection and open communication, effectively mimicking certain elements of human interaction. These findings reinforce self-disclosure theory, showing that chatbots capable of expressing emotion and understanding can build rapport and foster trust, leading to more meaningful engagement. Such insights position self-disclosure chatbots as promising tools for advancing educational support and mental well-being among students in higher education settings (Tardy & Smithson, 2018).

Additionally, this study highlights the theoretical value of integrating structured self-reflection frameworks, such as Koole et al.'s model (2011), into chatbot interactions. By guiding users through the stages of self-reflection, the SD chatbot enabled participants to gain an enhanced understanding of their stress-related thoughts and behaviors. The application of self-reflection frameworks within chatbot design enriches digital interventions, making them more effective at fostering introspection and insight among students. This study underscores that self-reflection frameworks can serve as a core component of chatbot-driven educational and therapeutic applications, adding both structure and depth that enhance the quality of interactions and support more nuanced learning outcomes.

Moreover, the study offers insights into how empathy and relatability in chatbots impact user engagement, self-reflection, and the accuracy of stress assessments. The SD chatbot's use of empathetic language and relatable content not only increased user engagement—evidenced by significantly longer session lengths and higher word counts—but also elicited more detailed and reflective responses. Notably, while the substantial increase in engagement enriched interactions, it did not result in a directly proportional improvement in assessment accuracy. This finding indicates that while self-disclosure fosters deeper and more extended interactions, it may also lead to the inclusion of tangential information that does not directly align with the SISCO-AS items. Nonetheless, the observed improvement in accuracy (ACC1) highlights the potential of self-disclosure to enhance the overall quality of chatbot-based assessments, underscoring the complex interplay between conversational depth and targeted data collection. These findings contribute to the theoretical understanding of emotional engagement in digital interactions, demonstrating that when chatbots create a supportive and comfortable environment, users are more willing to disclose personal information, thereby enriching the quality of collected data. Ultimately, this enhanced data quality can improve assessment outcomes and enable chatbots to deliver more precise and personalized support.

In summary, this study contributes to theoretical perspectives in educational technology by demonstrating that self-disclosure and empathy are essential elements in designing chatbots that promote meaningful engagement and effective self-reflection. By incorporating self-disclosure and self-reflection frameworks, chatbots in higher education can evolve from simple information-delivery tools to supportive partners in students' learning and well-being. These findings provide a foundational perspective for future research in educational technology, suggesting that self-disclosure and empathetic interaction in chatbot design can be further optimized to enhance educational outcomes and support mental health across various digital and educational settings.

Practical implication

The findings of this study provide valuable practical insights for enhancing chatbot-based stress management tools in educational technology, especially within higher education. By demonstrating that self-disclosure techniques—such as sharing relatable experiences and offering supportive responses—significantly improve user engagement and foster deeper self-reflection, this research offers actionable guidance for developers and educational institutions looking to create effective digital mental health tools. Incorporating self-disclosure into chatbot interactions not only makes chatbots more approachable but also strengthens their ability to capture complex emotional data, enhancing their effectiveness in both academic and emotional stress assessments.

Furthermore, the adaptive design of self-disclosure chatbots presents an opportunity for educational institutions to provide personalized support that adjusts to individual student needs. Rather than merely delivering generalized advice, chatbots with self-disclosure capabilities can dynamically adjust responses based on students' inputs, helping them explore tailored strategies for managing academic stress. This personalized support, embedded within a familiar digital platform, aligns well with the growing demand for accessible, student-centered educational technology that adapts to diverse learning and mental health needs.

Additionally, this study underscores the importance of humanizing AI-driven interventions to facilitate meaningful self-assessment. The application of self-disclosure and empathy in chatbots can extend beyond academic stress management to areas such as career counseling, behavioral coaching, and lifestyle guidance, where personal insights are equally essential. Developers and practitioners can apply these findings to design AI tools that provide richer, more engaging interactions, building trust and encouraging continuous use. Such chatbots not only improve the user experience but also help students develop an enhanced understanding of their emotions and behaviors, promoting personal growth and resilience.

In summary, the practical implications of this study point to a new paradigm in educational technology, where chatbots serve as empathetic companions that actively support students' learning and mental well-being. By incorporating self-disclosure, these chatbots can transform digital interactions into meaningful engagements that promote emotional resilience, reflective thinking, and sustainable growth. This approach can redefine the role of chatbots in higher education, positioning them as integral tools in both academic support and mental health initiatives.

Limitations and future directions

This study has several limitations that warrant further investigation. First, although the SD chatbot demonstrated a high ACC1 of 0.936, this result was based on data from only 50 students, all from the same university and nationality. The nature of students' interactions with the chatbot may vary depending on factors such as university environment and cultural background. Future research should validate these results across diverse educational and cultural contexts. Second, human researchers conducted the narrative analysis, enabling a nuanced understanding of self-disclosure in a sensitive domain but posing challenges in scalability. While generative AI tools offer promising alternatives, their ability to capture contextual nuances and emotional cues remains limited (Dentella et al., 2024; Farber, 2025). Ethical considerations also support the need for human oversight in handling personal narratives (Bouhouita-Guermech et al., 2023). Future studies should explore AI-driven methods to complement human analysis while ensuring ethical and interpretative accuracy. Third, our chatbot was developed using the GPT-4o model with prompt engineering. While effective, recent advancements in technology—such as retrieval-augmented generation and multi-agent systems—offer promising opportunities to enhance chatbot performance. Investigating these advancements may lead to more effective stress assessment tools. Lastly, while self-disclosure improved engagement and self-reflection, ethical concerns remain regarding chatbot authenticity in mental health applications. Some may view this as deceptive, particularly in sensitive mental health contexts, where trust and authenticity are crucial. Although our participants reported no issues, future research should examine perceptions of chatbot transparency and trustworthiness to ensure responsible implementation in sensitive contexts.

Despite these limitations, our study makes significant contributions by demonstrating that self-disclosure can enhance engagement, accuracy, and self-reflection in chatbot-based academic stress assessments. Our findings show that the SD chatbot encouraged deeper self-reflection, potentially leading to improved user engagement and more precise stress assessments. These insights not only advance the field of chatbot-based assessment but also underscore the potential of chatbots to support student well-being. By fostering more empathetic and reflective interactions, self-disclosure chatbots have the potential to improve educational outcomes and provide valuable mental health support in educational settings.

Acknowledgements
The authors would like to thank all participants for their great support and inspiration.

Author contributions
MP: conceptualization, methodology, investigation, writing—original draft, project administration; SF & KS: conceptualization, methodology, writing—review and editing, supervision, project administration. All authors read and approved the final manuscript.

Funding
This study was funded by Institute of Information & Communications Technology Planning & Evaluation (RS-2023-00262158).

Availability of data and materials
Details on the academic stress assessment questionnaire, as well as the GPT-4o API-based prompt engineering used for conversation protocols and stress evaluation methods, and example of our narrative analysis can be accessed at: https://osf.io/ch96q/?view_only=2f02931fbfe7482bae2b21fcbb5e9213.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 21 December 2024 Accepted: 7 April 2025
Published online: 09 May 2025

References
Ahlqvist-Björkroth, S., Thernström Blomqvist, Y., Nyberg, J., Normann, E., & Axelin, A. (2023). Improving NICU staff decision-making with parents in medical rounds: A pilot study of reflective group dialogue intervention. *Frontiers in Pediatrics, 11*, 1249345. https://doi.org/10.3389/fped.2023.1249345
Ahmed, A., Hassan, A., Aziz, S., Abd-Alrazaq, A. A., Ali, N., Alzubaidi, M., Al-Thani, D., Elhusein, B., Siddig, M. A., Ahmed, M., & Househ, M. (2023). Chatbot features for anxiety and depression: A scoping review. *Health Informatics Journal, 29*(1), 14604582221146720. https://doi.org/10.1177/14604582221146719
Alaoutinen, S. (2012). Evaluating the effect of learning style and student background on self-assessment accuracy. *Computer Science Education, 22*(2), 175–198. https://doi.org/10.1080/08993408.2012.692924
Altman, I. (1973). *Social penetration: The development of interpersonal relationships*. Rinehart, & Winston.
Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education, 17*(1), 3. https://doi.org/10.1186/s41239-020-0177-7
Arun, G., Perumal, V., Urias, F. P. J. B., Ler, Y. E., Tan, B. W. T., Vallabhajosyula, R., Tan, E., Ng, O., Ng, K. B., & Mogali, S. R. (2024). ChatGPT versus a customized AI chatbot (Anatbuddy) for anatomy education: A comparative pilot study. *Anatomical Sciences Education, 17*(7), 1396–1405. https://doi.org/10.1002/ase.2502
Atkins, S., & Murphy, K. (1993). Reflection: A review of the literature. *Journal of Advanced Nursing, 18*(8), 1188–1192. https://doi.org/10.1046/j.1365-2648.1993.18081188.x
Bamberg, M. (2012). Narrative analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, research designs: quantitative, qualitative, neuropsychological, and biological* (Vol. 2, pp. 85–102). American Psychological Association. https://doi.org/10.1037/13620-006
Barak, A., & Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior, 10*(3), 407–417. https://doi.org/10.1089/cpb.2006.9938
Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education, 20*(1), 62. https://doi.org/10.1186/s41239-023-00432-3
Booth, F., Potts, C., Bond, R., Mulvenna, M., Kostenius, C., Dhanapala, I., Vakaloudis, A., Cahill, B., Kuosmanen, L., & Ennis, E. (2023). A mental health and well-being chatbot: User event log analysis. *JMIR mHealth and uHealth, 11*, e43052. https://doi.org/10.2196/43052
Boud, D., Keogh, R., & Walker, D. (2013). *Reflection: Turning experience into learning*. Routledge.
Bouhouita-Guermech, S., Gogognon, P., & Bélisle-Pipon, J. C. (2023). Specific challenges posed by artificial intelligence in research ethics. *Frontiers in Artificial Intelligence, 6*, 1149082. https://doi.org/10.3389/frai.2023.1149082
Bourner, T. (2003). Assessing reflective learning. *Education + Training, 45*(5), 267–272. https://doi.org/10.1108/00400910310484321
Cha, S., Loeser, M., & Seo, K. (2024). The impact of AI-based course-recommender system on students' course-selection decision-making process. *Applied Sciences, 14*(9), 3672. https://doi.org/10.3390/app14093672
Choi, D. S., Park, J., Loeser, M., & Seo, K. (2024). Improving counseling effectiveness with virtual counselors through nonverbal compassion involving eye contact, facial mimicry, and head-nodding. *Scientific Reports, 14*(1), 506. https://doi.org/10.1038/s41598-023-51115-y
Chung, L. L., & Kang, J. (2023). I'm hurt too: The effect of a chatbot\s reciprocal self-disclosures on users' painful experiences. *Archives of Design Research, 36*(4), 67–84. https://doi.org/10.15187/adr.2023.11.36.4.67
Cohen, S. (1997). *Measuring stress: A guide for health and social scientists*. Oxford University Press.
Croes, E., Antheunis, M. L., & He, L. (2023, September). *You go first: The effects of self-disclosure reciprocity in human-chatbot interactions*. In 2023 11th international conference on affective computing and intelligent interaction workshops and demos (ACIIW) (pp. 1–4). IEEE. https://doi.org/10.1109/ACIIW59127.2023.10388091
Cui, Y., Lee, Y. J., Jamieson, J., Yamashita, N., & Lee, Y. C. (2024). Exploring effects of chatbot's interpretation and self-disclosure on mental illness stigma. *Proceedings of the ACM on Human-Computer Interaction, 8*(CSCW1), 1–33. https://doi.org/10.1145/3637329
Degani, A., Goldman, C. V., Deutsch, O., & Tsimhoni, O. (2017). On human–machine relations. *Cognition, Technology & Work, 19*(2), 211–231. https://doi.org/10.1007/s10111-017-0417-3
Dentella, V., Günther, F., Murphy, E., Marcus, G., & Leivada, E. (2024). Testing AI on language comprehension tasks reveals insensitivity to underlying meaning. *Scientific Reports, 14*(1), 28083. https://doi.org/10.1038/s41598-024-79531-8
Farber, S. (2025). Comparing human and AI expertise in the academic peer review process: towards a hybrid approach. *Higher Education Research & Development*. https://doi.org/10.1080/07294360.2024.2445575
Gaudette, L., spsampsps Japkowicz, N. (2009). *Evaluation methods for ordinal classification*. In Advances in artificial intelligence: 22nd Canadian conference on artificial intelligence, Canadian AI 2009 Kelowna, Canada, May 25–27, 2009 Proceedings 22 (pp. 207–210). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01818-3_25
Ghosh, S., Mitra, B., & De, P. (2020, April). *Towards improving emotion self-report collection using self-reflection*. In extended abstracts of the 2020 CHI conference on human factors in computing systems (pp. 1–8). https://doi.org/10.1145/3334480.3383019

Gibbs, G. (1988). *Learning by doing: A guide to teaching and learning methods. Further education unit*. Oxford Polytechnic.

Gladding, S. T. (2009). *Counseling: A comprehensive profession, 6/E*. Pearson Education India.

Goddard, Q., Moton, N., Hudson, J., & He, H. A. (2024, May). *A chatbot won't judge me: An exploratory study of self-disclosing chatbots in introductory computer science classes*. In Proceedings of the 26th Western Canadian conference on computing education (pp. 1–7). https://doi.org/10.1145/3660650.3660662

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: An International Journal, 30*(8), 821–835. https://doi.org/10.2224/sbp.2002.30.8.821

Habicht, J., Viswanathan, S., Carrington, B., Hauser, T. U., Harper, R., & Rollwage, M. (2024). Closing the accessibility gap to mental health treatment with a personalized self-referral Chatbot. *Nature Medicine*. https://doi.org/10.1038/s41591-023-02766-x

Hanley, T., Ersahin, Z., Sefi, A., & Hebron, J. (2017). Comparing online and face-to-face student counselling: What therapeutic goals are identified and what are the implications for educational providers? *Journal of Psychologists and Counsellors in Schools, 27*(1), 37–54. https://doi.org/10.1017/jgc.2016.20

Haque, M. R., & Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR mHealth and uHealth, 11*(1), e44838. https://doi.org/10.2196/44838

Hjeltnes, A., Binder, P. E., Moltu, C., & Dundas, I. (2015). Facing the fear of failure: An explorative qualitative study of client experiences in a mindfulness-based stress reduction program for university students with academic evaluation anxiety. *International Journal of Qualitative Studies on Health and Well-Being, 10*(1), 27990. https://doi.org/10.3402/qhw.v10.27990

Idowu, J. A., Koshiyama, A. S., & Treleaven, P. (2024). Investigating algorithmic bias in student progress monitoring. *Computers and Education: Artificial Intelligence, 7*, 100267. https://doi.org/10.1016/j.caeai.2024.100267

Ignatius, E., & Kokkonen, M. (2007). Factors contributing to verbal self-disclosure. *Nordic Psychology, 59*(4), 362–391. https://doi.org/10.1027/1901-2276.59.4.362

Jiménez-Mijangos, L. P., Rodríguez-Arce, J., Martínez-Méndez, R., & Reyes-Lagos, J. J. (2023). Advances and challenges in the detection of academic stress and anxiety in the classroom: A literature review and recommendations. *Education and Information Technologies, 28*(4), 3637–3666. https://doi.org/10.1007/s10639-022-11324-w

Jin, S. H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education, 20*(1), 37. https://doi.org/10.1186/s41239-023-00406-5

Jin, Y., Chen, L., Zhao, X., & Cai, W. (2024). The way you assess matters: User interaction design of survey chatbots for mental health. *International Journal of Human-Computer Studies, 189*, 103290. https://doi.org/10.1016/j.ijhcs.2024.103290

Jo, E., Jeong, Y., Park, S., Epstein, D. A., & Kim, Y. H. (2024, May). *Understanding the impact of long-term memory on self-disclosure with large language model-driven chatbots for public health intervention*. In Proceedings of the CHI conference on human factors in computing systems (pp. 1–21). https://doi.org/10.1145/3613904.3642420

Kang, E., & Kang, Y. A. (2024). Counseling chatbot design: The effect of anthropomorphic chatbot characteristics on user self-disclosure and companionship. *International Journal of Human-Computer Interaction, 40*(11), 2781–2795. https://doi.org/10.1080/10447318.2022.2163775

Kim, Y. (2024). AI health counselor's self-disclosure: its impact on user responses based on individual and cultural variation. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2024.2316373

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Koole, S., Dornan, T., Aper, L., Scherpbier, A., Valcke, M., Cohen-Schotanus, J., & Derese, A. (2011). Factors confounding the assessment of reflection: A critical review. *BMC Medical Education, 11*, 1–9. https://doi.org/10.1186/1472-6920-11-104

Korthagen, F., & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching, 11*(1), 47–71. https://doi.org/10.1080/1354060042000337093

Lee, J., Lee, D., & Lee, J. G. (2024). Influence of rapport and social presence with an AI psychotherapy chatbot on users' self-disclosure. *International Journal of Human-Computer Interaction, 40*(7), 1620–1631. https://doi.org/10.1080/10447318.2022.2146227

Lee, Y. C., Cui, Y., Jamieson, J., Fu, W., & Yamashita, N. (2023, April). *Exploring effects of chatbot-based social contact on reducing mental illness stigma*. In Proceedings of the 2023 CHI conference on human factors in computing systems (pp. 1–16). https://doi.org/10.1145/3544548.3581384

Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020, April). *"I hear you, I feel you": encouraging deep self-disclosure through a chatbot*. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1–12). https://doi.org/10.1145/3313831.3376175

Liang, K. H., Shi, W., Oh, Y. J., Wang, H. C., Zhang, J., & Yu, Z. (2024). *Dialoging resonance in human-chatbot conversation: How users perceive and reciprocate recommendation chatbot's self-disclosure strategy*. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), (pp. 1–28). https://doi.org/10.1145/3653691

Lundgren, M. (2024). Large language models in student assessment: Comparing ChatGPT and human graders. arXiv preprint arXiv:2406.16510. https://doi.org/10.48550/arXiv.2406.16510

Macías, A. B. (2006). Un modelo conceptual para el estudio del estrés académico. *Revista Electrónica De Psicología Iztacala, 9*(3), 110–129.

Macías, A. B. (2007). El Inventario SISCO del estrés académico. *Investigación Educativa Duranguense, 7*, 90–93.

Mamede, S., & Schmidt, H. G. (2004). The structure of reflective practice in medicine. *Medical Education, 38*(12), 1302–1308.

Mead, J. (2019). *What session length can tell you about your chatbot performance*. Inf. Commun

Neijens, P., Araujo, T., Möller, J., de Vreese, C. H., Segijn, C. M., Vraga, E., Hopp, F. R., & Bakker, B. N. (2024). Self-report measures. *Measuring exposure and attention to media and communication: Solutions to wicked problems* (pp. 41–86). Amsterdam University Press. https://doi.org/10.2307/jj.17610809.8

Park, G., Chung, J., & Lee, S. (2023). Effect of AI chatbot emotional disclosure on user satisfaction and reuse intention for mental health counseling: A serial mediation model. *Current Psychology, 42*(32), 28663–28673. https://doi.org/10.1007/s12144-022-03932-z

Park, J. I., Abbasian, M., Azimi, I., Bounds, D., Jun, A., Han, J., McCarron, R.M., Borelli, J., Safavi, P., Mirbaha, S., Li, J., Mahmoudi, M., Wiedenhoeft, C., & Rahmani, A.M. (2024). Building trust in mental health chatbots: Safety metrics and LLM-based evaluation tools. arXiv preprint arXiv:2408.04650. https://doi.org/10.48550/arXiv.2408.04650

Park, M., & Seo, K. (2023). *Passive vs. Active: Which behavior is better for predicting at-risk students in online learning?*. Proceedings of HCI Korea. p. 815–821.

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. *International Journal of Adolescence and Youth, 25*(1), 104–112. https://doi.org/10.1080/02673843.2019.1596823

Petzel, Z. W., & Sowerby, L. (2025). Prejudiced interactions with large language models (LLMs) reduce trustworthiness and behavioral intentions among members of stigmatized groups. *Computers in Human Behavior, 165*, 108563. https://doi.org/10.1016/j.chb.2025.108563

Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development, 80*(4), 1016–1038. https://doi.org/10.1111/j.1467-8624.2009.01314.x

Schick, A., Feine, J., Morana, S., Maedche, A., & Reininghaus, U. (2022). Validity of chatbot use for mental health assessment: Experimental study. *JMIR mHealth and uHealth, 10*(10), e28082. https://doi.org/10.2196/28082

Schillings, C., Meißner, E., Erb, B., Bendig, E., Schultchen, D., & Pollatos, O. (2024). Effects of a chatbot-based intervention on stress and health-related parameters in a stressed sample: Randomized controlled trial. *JMIR Mental Health, 11*(1), e50454. https://doi.org/10.2196/50454

Seo, K., Dodson, S., Harandi, N. M., Roberson, N., Fels, S., & Roll, I. (2021a). Active learning with online video: The impact of learning context on engagement. *Computers & Education, 165*, 104132. https://doi.org/10.1016/j.compedu.2021.104132

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021b). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*, 1–23. https://doi.org/10.1186/s41239-021-00292-9

Seo, K., Yoo, M., Dodson, S., & Jin, S. H. (2024). Augmented teachers: K–12 teachers' needs for artificial intelligence's complementary role in personalized learning. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2024.2330525

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Song, C., Park, M., Kim, D., & Seo, K. (2024). *ChatGPT as an assistant of human teacher: Student competency analysis through prompt engineering*. Proceedings of HCI Korea. pp. 175–182.

Steare, T., Muñoz, C. G., Sullivan, A., & Lewis, G. (2023). The association between academic pressure and adolescent mental health problems: A systematic review. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2023.07.028

Stockhausen, L. (1994). The clinical learning spiral: A model to develop reflective practitioners. *Nurse Education Today, 14*(5), 363–371. https://doi.org/10.1016/0260-6917(94)90031-0

Sung, G., Bhinder, H., Feng, T., & Schneider, B. (2023). Stressed or engaged? Addressing the mixed significance of physiological activity during constructivist learning. *Computers & Education, 199*, 104784. https://doi.org/10.1016/j.compedu.2023.104784

Tardy, C. H., & Smithson, J. (2018). *Self-disclosure: Strategic revelation of information in personal and professional relationships 1*. The handbook of communication skills. pp. 217–258.

Toulme, P., Nanaw, J., & Apostolellis, P. M. (2023). *A Chatbot for depression screening based on the PHQ-9 assessment*. The sixteenth international conference on advances in computer-human interactions (ACHI). pp. 97–105.

Travers, C. J., Morisano, D., & Locke, E. A. (2015). Self-reflection, growth goals, and academic outcomes: A qualitative study. *British Journal of Educational Psychology, 85*(2), 224–241. https://doi.org/10.1111/bjep.12059

Tsumura, T., & Yamada, S. (2023). Influence of agent's self-disclosure on human empathy. *PLoS ONE, 18*(5), e0283955. https://doi.org/10.1371/journal.pone.0283955

Ulrich, S., Lienhard, N., Künzli, H., & Kowatsch, T. (2024). A chatbot-delivered stress management coaching for students (MISHA app): Pilot randomized controlled trial. *JMIR mHealth and uHealth, 12*, e54945. https://doi.org/10.2196/54945

van der Schyff, E. L., Ridout, B., Amon, K. L., Forsyth, R., & Campbell, A. J. (2023). Providing self-led mental health support through an artificial intelligence-powered chat bot (Leora) to meet the demand of mental health care. *Journal of Medical Internet Research, 25*, e46448. https://doi.org/10.2196/46448

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research, 13*(3), 253–266. https://doi.org/10.1177/1094670510375599

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*(1), 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction (TOCHI), 27*(3), 1–37. https://doi.org/10.1145/3381804

Xue, J., Wang, Y. C., Wei, C., Liu, X., Woo, J., & Kuo, C. C. J. (2023). Bias and fairness in chatbots: An overview. arXiv preprint arXiv:2309.08836. https://doi.org/10.48550/arXiv.2309.08836

Zhang, S., Zhao, X., Nan, D., & Kim, J. H. (2024). Beyond learning with cold machine: Interpersonal communication skills as anthropomorphic cue of AI instructor. *International Journal of Educational Technology in Higher Education, 21*(1), 27. https://doi.org/10.1186/s41239-024-00465-2

Publisher's Note