University of Tartu
Faculty of Social Sciences
Institute of Education
Curriculum of Educational Technology

André von Eye Corleta

Supporting Cognitive Test Anxiety in Self-Regulated Learning: Comparing AI and Human Feedback in Digital Test Preparation

MA thesis

Supervisor: Prof. Natalia Edisherashvili

Tartu, 2026

Abstract

Cognitive test anxiety represents one of the most consequential emotional barriers in high-stakes educational contexts, yet its regulation through digital platform design remains underexplored. Grounded in Self-Regulated Learning theory and Control-Value Theory, this study examined whether elaborated feedback delivered by an AI system or a human tutor can attenuate cognitive test anxiety among learners preparing for a competitive university entrance examination in Brazil, and whether feedback source and modality matter to emotional outcomes. A quasi-experimental pre-post design was employed within a digital test-preparation platform, with 41 participants assigned to three groups: AI feedback, human feedback, and a control control. The Cognitive Test Anxiety Scale, Second Edition (CTAS-2) was administered before and after the three-week intervention. A convergent mixed-methods design integrated quantitative findings with qualitative data from semi-structured interviews conducted with five participants. No within-group changes reached statistical significance in any condition. However, a significant between-group difference emerged ($H = 6.80$, $p = .033$), driven by markedly divergent trajectories: the AI-feedback group showed a non-significant trend toward anxiety reduction ($r = .43$), while the human-feedback group showed an increase ($r = .61$). Qualitative analysis yielded three themes: AI as the default pragmatic tool for active help-seeking, the irreducible emotional value of human presence, and passive feedback as a motivational catalyst anchored in goal-setting. The findings support a functional differentiation of feedback roles by source and modality, with design implications for emotionally responsive digital learning platforms.

Keywords: cognitive test anxiety, elaborated feedback, AI feedback, self-regulated learning, Control-Value Theory, digital test preparation

1. Introduction

Learning is not a purely cognitive act. Students do not merely process information: they experience the joy of discovering something new, worry about falling short, and sometimes doubt their own capacity entirely. The architecture of digital learning amplifies this challenge: where traditional classroom formats once structured the pace, sequence, and social texture of education, online environments transfer that responsibility almost entirely to the individual (Broadbent & Poon, 2015). This redistribution of responsibility places self-regulated learning (SRL) and, crucially, emotional regulation, at the center of educational success in high-stakes digital contexts.

Building on foundational work by Zimmerman (1989), SRL is typically conceptualized as a cyclical process comprising forethought, performance, and self-reflection, with each phase engaging intertwined cognitive, metacognitive, motivational, and emotional dimensions. In exam-preparation contexts, the emotional dimension acquires particular salience: sustained evaluative pressure generates the achievement emotions that shape learner engagement and performance (Pekrun et al., 2007). Among these, cognitive test anxiety, a pattern of task-irrelevant worry that interferes with learning not only during testing but across the entire preparation cycle, occupies a central position (Cassady, 2004; Cassady & Johnson, 2002). Control-Value Theory (CVT) provides the theoretical account of how it develops: anxiety emerges when high task value meets diminished perceived control (Pekrun et al., 2007). Understanding what sustains or disrupts this appraisal pattern is the central concern of the present study.

In parallel, as educational technology evolves, artificial intelligence (AI) has become an established presence across diverse instructional settings. AI-driven systems are now embedded within online learning platforms, higher education curricula, and select K-12 environments (Banihashem et al., 2025). The integration of AI into learning environments primarily supports personalization and data analysis, with key applications ranging from tutoring and real-time feedback to risk prediction and student profiling (Banihashem et al., 2025; Edisherashvili et al., 2025; Prasse et al., 2024; Xia et al., 2026). One pedagogical mechanism with particular promise in this regard is elaborated feedback. Feedback has long been established as one of the most influential factors in student learning, with meta-analytic evidence positioning its average effect on performance among the strongest documented in educational research (Hattie & Timperley, 2007; Wisniewski et al., 2020). In digital learning environments, feedback can be delivered either by human instructors or by AI-driven systems. Yet despite the rapid expansion of AI feedback systems in education, their capacity to support learners' emotional regulation, and how this compares to human-delivered feedback, remains unexamined (Banihashem et al., 2025).

The convergence of heightened emotional demands in high-stakes digital learning contexts and the rapid expansion of AI-based instructional tools form the backdrop for the present study. The central question it addresses is whether elaborated feedback delivered by AI or by a human tutor can attenuate cognitive test anxiety among learners preparing for competitive university entrance examinations in Brazil, and whether the source of that feedback matters to the emotional outcomes learners experience.

1.1 Research Gap

The first gap is conceptual. While SRL research acknowledges emotional regulation as a core dimension, studies explicitly targeting it, and cognitive test anxiety in particular, as an outcome variable in digital learning contexts remain scarce (Banihashem et al., 2025; Edisherashvili et al., 2022). Fewer still examine whether and how the pedagogical and technological design of platforms can actively attenuate it.

Moreover, a second gap is demographic. Reviews of SRL and AI-SRL research reveal a strong concentration in higher education (Banihashem et al., 2025). Learners preparing for competitive university entrance examinations occupy a transitional position not captured by either higher education or school-based research, facing sustained high-stakes emotional demands and potentially distinctive self-regulatory profiles (Seddigh et al., 2016; Wu et al., 2022).

Finally, while emerging research has begun to compare AI-based and human-based feedback on performance and motivation (Banihashem et al., 2026; Yildiz Durak & Onan, 2025), no study has examined whether these modalities differ in their capacity to reduce cognitive test anxiety specifically, in authentic high-stakes test-preparation contexts. It is not yet known whether the source of feedback, artificial or human, matters to the emotional outcomes learners experience.

Against this backdrop, the present study positions itself at the intersection of three converging gaps: emotional, demographic, and technological, within digital test-preparation environments. The study generates evidence on whether and how elaborated feedback, delivered by AI or by a human tutor, can attenuate cognitive test anxiety among learners navigating one of the most high-stakes educational transitions in the Brazilian system.

2. Literature Review

This review develops the theoretical foundations of the present study across four interconnected areas: self-regulated learning as the overarching framework (Section 2.1); the emotional dimension of SRL, including achievement emotions, Control-Value Theory, and cognitive test anxiety (Section 2.2); feedback as the primary pedagogical intervention and its role in emotion regulation (Section 2.3); and the intersection of online learning and AI, including how AI-generated feedback compares to human feedback as a source of emotionally relevant instructional support (Section 2.4).

2.1. Self-Regulated Learning

Self-regulated learning (SRL) is defined as a goal-oriented process through which individuals actively manage their study to achieve specific learning objectives (Winne & Hadwin, 1998; Zimmerman, 1989). Zimmerman's foundational cyclical model frames this as a three-phase process of forethought (planning), performance (execution), and self-reflection (appraisal), in which learners continuously plan, monitor, and adapt their strategies in pursuit of their goals.

Subsequent theoretical work progressively expanded beyond this cognitive foundation (Panadero, 2017). Pintrich (2000) explicitly integrated motivational processes, while Boekaerts (1996) foregrounded affect through a dual-processing model balancing mastery goals with well-being regulation. Efklides (2011) further distinguished macro-level self-regulation from micro-level metacognitive experiences, emphasizing affective monitoring as a continuous process. Taken together, this theoretical evolution reflects a growing consensus that SRL cannot be adequately understood through cognitive and metacognitive processes alone: motivational and emotional regulation are constitutive dimensions, not peripheral additions (Panadero, 2017). Therefore, the SRL construct is understood as a multifaceted process characterized by four dimensions: cognitive, metacognitive, emotional, and motivational (Boekaerts, 1996; Efklides, 2011; Pintrich, 1995).

The empirical significance of this expanded understanding is well established. Strong self-regulation skills consistently predict improved academic performance). Crucially, this evidence extends to the motivational and emotional dimensions: students who effectively regulate their affective states, including anxiety, demonstrate greater persistence, more adaptive strategy use, and stronger performance outcomes (Chen et al., 2024). It is this emotional dimension of SRL, and the conditions under which it is most at risk, that the sections that follow address.

2.2 Achievement Emotions Regulation and Test Anxiety within SRL

Emotions are pervasive in academic contexts, influencing memory, cognition, and learner persistence (Du et al., 2023; Tyng et al., 2017). Within the SRL literature, they are conceptualized as integral to motivation, regulation, and academic engagement (Edisherashvili et al., 2025).

Achievement emotions are defined as those directly linked to achievement activities, such as studying, or their outcomes (Pekrun et al., 2007). They are characterized by a specific valence (positive or negative) and a distinct object focus (activity-based or outcome-based). Among the negative outcome emotions, test anxiety occupies a central position: it is the most extensively studied achievement emotion in educational research (Pekrun et al., 2007), reflecting its capacity to disrupt the cognitive, motivational, and self-regulatory processes on which academic performance depends.

Test anxiety is a multidimensional construct, encompassing a cognitive component (worry and task-irrelevant rumination), and a somatic component (physiological arousal) (Liebert & Morris, 1967; Mandler & Sarason, 1952). The cognitive dimension has been consistently identified as the stronger predictor of academic performance (Cassady & Johnson, 2002): it generates intrusive thoughts that impair knowledge acquisition not only during testing but across the entire learning-testing cycle, including in low-stakes conditions (Cassady, 2004; Cassady & Johnson, 2002; Wine, 1971; Tobias, 1983). Meta-analytic evidence places its association with academic performance in the small to moderate range (von der Embse et al., 2018), though effects are consistently negative and particularly pronounced in high-stakes entrance examinations contexts (Karatas et al., 2013).

The theoretical account of how test anxiety develops is provided by Control-Value Theory (CVT; Pekrun et al., 2007): anxiety arises when high task value meets low perceived control, a configuration structurally present in exam-preparation contexts (Pekrun et al., 2017). CVT is particularly suited to the present study, because elaborated feedback acts directly on perceived control: it confirms and reframes progress, raising the learner's sense of agency (Hattie & Timperley, 2007). This mechanism is examined in detail in Section 2.3.

2.3 Feedback for learning and anxiety regulation

Feedback is consistently identified as one of the most influential factors in student learning, yet its effects vary substantially depending on how it is designed and delivered). A dimension that the feedback literature has relatively underexplored concerns the direction of initiation: whether feedback is delivered to the learner by an external agent (passive feedback) or actively sought by the learner themselves (active feedback) (Butler & Winne, 1995). The sections that follow address these questions through two lenses: the foundational theoretical model specifying what feedback does and at which level it operates (Section 2.3.1), and the mechanisms through which feedback acts on achievement emotion regulation and help-seeking (Section 2.3.2).

2.3.1 Classic Feedback Theory

The most widely cited theoretical account of feedback in educational contexts is offered by Hattie and Timperley (2007), whose model frames feedback as information provided by an agent regarding aspects of one's performance or understanding, with the primary purpose of reducing the discrepancy between a learner's current state and a desired goal. This discrepancy-reduction logic organizes effective feedback around three interconnected questions: Where am I going? (feed up: clarifying goals and success criteria), How am I going? (feed back: providing information about progress relative to those goals), and Where to next? (feed forward: directing future learning strategies and effort) (Hattie & Timperley, 2007).

A central contribution of the model is its differentiation of four levels at which feedback can operate. Task-level feedback addresses the correctness or quality of a specific response. Process-level feedback targets the strategies and cognitive operations underlying task performance. Self-regulation feedback supports learners' monitoring and regulation of their own learning processes, including self-efficacy, help-seeking, and error detection. Finally, self-level feedback addresses personal attributes, typically in the form of praise or criticism directed at the learner rather than the work. Hattie and Timperley (2007) argue that feedback is most effective when directed at the process and self-regulation levels, as these engage deeper learning and foster independent strategy use, whereas self-level feedback rarely translates into learning gains. Feedback at these levels, providing not only information about current performance but also orientation toward future action, is collectively referred to as elaborated feedback (Hattie & Timperley, 2007; Wisniewski et al., 2020).

Meta-analytic evidence confirms both the power of feedback and the importance of its form. A synthesis of 12 meta-analyses places the average effect of feedback on student achievement among the strongest instructional influences documented in educational research (Hattie & Timperley, 2007). Wisniewski et al. (2020) demonstrate that effect sizes vary substantially by feedback type: high-information feedback, which provides guidance at task and self-regulation levels, produces effects nearly twice as large as corrective feedback alone, while reinforcement and praise show markedly weaker effects. Critically for the present study, feedback exerts stronger effects on cognitive outcomes than on motivational or behavioral ones, a distinction directly relevant to a study targeting the cognitive dimension of test anxiety.

2.3.2 Feedback and Emotional Regulation

While Hattie and Timperley's (2007) model establishes what feedback does at the cognitive level, a fuller account of its role in learning requires attending to the affective processes it simultaneously activates. Butler and Winne (1995) argued that feedback is not merely an external input but is structurally embedded in self-regulated learning: as learners monitor progress, internal feedback continuously generates affect: negative when progress falls below expectations, positive when it meets or exceeds them. External feedback intervenes in this cycle by recalibrating the learner's appraisal of the gap between current state and goal. This mechanism maps directly onto Pekrun's (2007) Control-Value Theory: feedback that confirms progress or reframes failure acts on perceived control, a key determinant of negative achievement emotions including anxiety. Elaborated feedback, which orients the learner toward future action rather than merely reporting outcomes, is therefore theorised to attenuate cognitive test anxiety by restoring a sense of agency over the path to the goal.

Meta-analytic evidence indicates that negative feedback substantially reduces perceived competence even when intrinsic motivation remains intact (Fong et al., 2019), and perceived competence is precisely the variable through which CVT predicts anxious arousal. Crucially, however, the same evidence shows that negative feedback containing instructional guidance (directions for improvement) substantially attenuates this effect, consistent with Hattie and Timperley's (2007) argument that feed-forward is the most impactful dimension of effective feedback. Despite the theoretical coherence of this chain, direct empirical research on the feedback–test anxiety relationship remains sparse.

A dimension of the feedback literature with direct implications for emotional regulation concerns the role of learner agency in initiating feedback itself (Butler & Winne, 1995). A theoretically meaningful contrast can be drawn between feedback that is externally initiated (passive feedback) and feedback the learner actively solicits (active feedback), what the SRL literature terms help-seeking: a social strategy through which learners request the specific support they need to advance in a task (Martín-Arbós et al., 2021). Help-seeking is therefore not merely a precursor to receiving feedback; it is itself a self-regulatory act that signals the learner's perceived agency. The emotional implications are substantial: Edisherashvili et al. (2025) found that help-seeking strategies produced one of the largest effect sizes in a systematic review of achievement emotions in online adult learning, with a very large positive impact on hope and relief, and peer help-seeking was associated with significant anxiety reduction. These findings suggest that the act of requesting feedback, not merely the quality of the response received, carries intrinsic emotional value by restoring perceived control. Passive feedback, however well-designed, does not activate this self-regulatory pathway.

This distinction between active and passive feedback is consequential for the present study, which investigates learner-initiated and externally delivered feedback as distinct pathways to emotional regulation: it foregrounds the question whether AI-generated feedback can function as a medium for active feedback, or whether its typical delivery structure positions learners as passive recipients. This question is taken up in Section 2.4.

2.4 Online SRL and Artificial Intelligence

The theoretical case developed in Sections 2.2 and 2.3 acquires particular urgency in online learning environments, where platforms transfer responsibilities almost entirely to the learner, amplifying both the self-regulatory demands and the emotional risks that accompany them (Broadbent & Poon, 2015; Jin et al., 2023). It is against this backdrop that AI has emerged as a significant presence in digital education.

2.4.1 Online SRL and AI as Instructional Support

The autonomy that defines online learning is both its primary affordance and its chief pedagogical risk. Research consistently documents that many learners lack the self-regulatory skills to exploit this autonomy productively: motivation declines in the absence of external accountability (Du et al., 2023), monitoring is infrequent and unreliable (Zimmerman, 2002), and feedback is underutilized as a regulatory resource (Broadbent & Poon, 2015).

Artificial intelligence has become the primary mechanism through which this support is increasingly operationalized in digital education. Such systems are now embedded across online learning platforms, higher education curricula, and K-12 settings, with core capabilities including adaptivity, intelligent feedback, and automated social presence (Banihashem et al., 2025; Edisherashvili et al., 2025; Stalmach et al., 2025). In the domain of SRL specifically, AI applications range from goal-setting organizers and adaptive study planners to intelligent tutoring systems capable of providing real-time regulatory support (Jin et al., 2023; Molenaar et al., 2023) . A systematic mapping study by Banihashem et al. (2025) documents the rapid expansion of this field while identifying emotional regulation as a promising but underexplored application, a gap the present study addresses directly.

2.4.2 AI for Achievement Emotions and Test Anxiety Regulation

Evidence on AI's capacity to regulate emotions in learning contexts is promising but temporally bounded. Meta-analytic findings indicate that AI-based conversational agents can produce meaningful short-term reductions in psychological distress and depression (Sohn et al., 2026), a pattern corroborated by an earlier systematic review reporting significant reductions in depressive symptoms across randomized trials (Abd-Alrazaq et al., 2020). The same evidence base, however, consistently fails to demonstrate sustained effects on broader well-being indicators, with the asymmetry attributed to well-being's greater stability as a construct, one that requires prolonged engagement to shift, a condition brief AI interactions rarely meet (Sohn et al., 2026). Pavlopoulos et al. (2024) further caution that dependence on AI systems without sustained human oversight may, over time, erode rather than consolidate emotional self-regulation capacities.

Direct empirical evidence on AI's capacity to reduce test anxiety specifically remains sparse. A recent randomized controlled trial examined the impact of AI-generated study material and feedback on pharmacy students' anxiety and performance in high-stakes OSCEs (Ali et al., 2025). No significant difference in test anxiety was found between intervention and control groups, suggesting that access to generative AI tools alone may be insufficient to attenuate evaluative anxiety, with effectiveness likely depending on how AI is integrated and tailored to students' needs, a consideration central to the elaborated feedback design examined here.

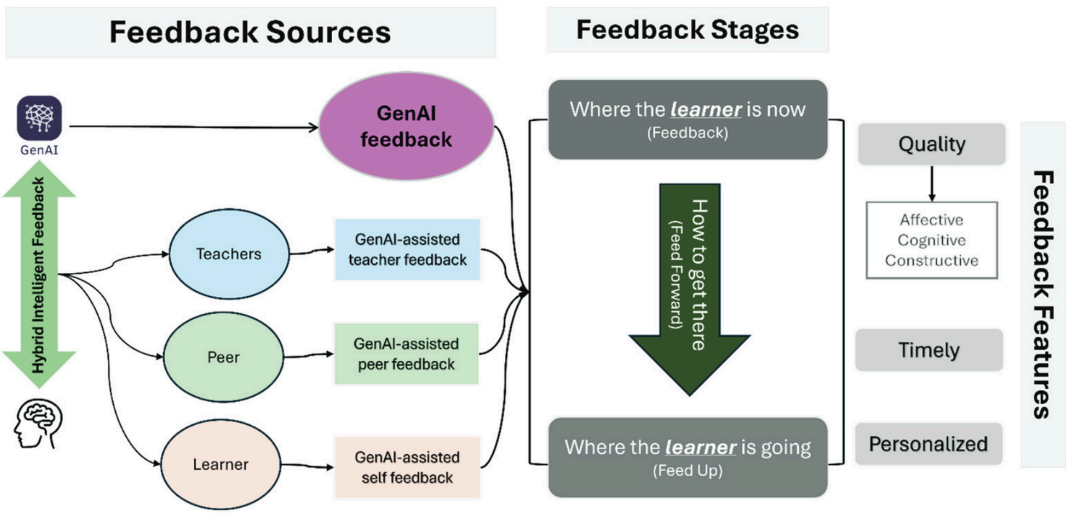
This temporal ceiling raises a broader theoretical question: whether short-term anxiety reduction during test preparation is, in fact, the optimal outcome to target. Stress inoculation training (SIT) proposes that resilience to future stressors is built through graduated, controlled exposure to stress conditions approximating the target environment, not through their elimination (Meichenbaum & Deffenbacher, 1988; Saunders & Driskell, 1996). This is corroborated by the concept of hormesis, which holds that moderate stress exposure can enhance performance and adaptive capacity, whereas both insufficient and excessive stress produce suboptimal outcomes (Hill et al., 2024). Applied to test-preparation contexts, interventions that reduce anxiety during preparation may inadvertently deprive learners of the evaluative arousal necessary to build regulatory resources for the exam itself. This distinction between anxiety reduction as a proximate goal and regulatory capacity as a distal one is returned to in the discussion.

2.4.3 Artificial Intelligence as a Feedback Source

Generative AI has opened new possibilities for feedback provision in education. Banihashem et al. (2026) position GenAI feedback within Hattie and Timperley's (2007) feedback model: it is capable of addressing all three core questions of the feedback process (feed back, feed up, feed forward), functioning not merely as a supplement to human feedback sources but as a distinct instructional role depending on pedagogical context. Feedback source structure, alongside the feedback stages and features is represented in Figure 1.

Figure 1

Human Gen-AI collaborates with humans, but stages and features remain the same



Note. Extracted from (Banihashem et al., 2026).

Empirical evidence supports the viability of this role while identifying its limits. Studies comparing ChatGPT-generated and peer-generated feedback suggest the two are complementary: AI feedback tends to be more descriptive and style-focused, while peer feedback more frequently addresses specific content issues (Banihashem et al., 2026). In tasks such as essay writing, learning outcomes with GenAI feedback are comparable to those achieved with human feedback, with no significant difference in student preference (Banihashem et al., 2026). A systematic review by Yildiz Durak and Onan (2025) confirms positive effects of AI-based feedback systems on motivation, self-regulation, and academic performance, particularly in higher education and language learning, while cautioning against insufficient contextual sensitivity, algorithmic bias, and a potential decline in human interaction.

2.5 Present Study Aim and Research Questions

The theoretical and empirical threads developed across this review converge on a productive tension. Cognitive test anxiety emerges when high task value meets diminished perceived control (Pekrun et al., 2007), and elaborated feedback is one of the most powerful mechanisms available to intervene in precisely this dynamic. Yet the conditions under which AI-generated feedback can produce this regulatory effect, and whether it does so comparably to human-delivered feedback, remain empirically unresolved. Existing studies either examine emotional outcomes as secondary to performance, address general anxiety rather than cognitive test anxiety specifically, or lack the direct AI-versus-human comparison that would allow design-relevant conclusions to be drawn (Ali et al., 2025; Banihashem et al., 2025).

This study examines how AI-based and human-provided elaborated feedback influence students' cognitive test anxiety in a digital test-preparation context, and aims to derive design-relevant insights for emotionally supportive learning environments. It is structured around three research questions:

RQ1. To what extent does the provision of elaborated feedback influence changes in students' cognitive test anxiety during a three-week online test-preparation program?

RQ2. To what extent do AI-based and human-based feedback modalities differ in their effects on students' cognitive test anxiety during a three-week online test-preparation program?

RQ3. How do learners perceive the role of AI- and human-based feedback, across active and passive modalities, in influencing their cognitive test anxiety during exam preparation?

3. Method

This study investigates whether elaborated feedback delivered by either an AI system or a human tutor can reduce cognitive test anxiety among students preparing for a high-stakes university entrance exam in Brazil.

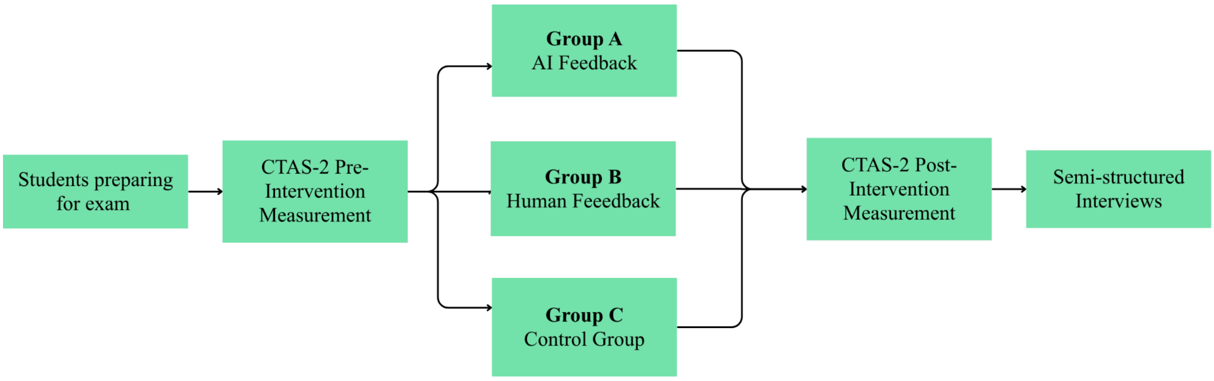
3.1 Study Design

The study employed a quasi-experimental pre-post design within a proprietary digital test-preparation platform (A2X Med), using the Cognitive Test Anxiety Scale, Second Edition (CTAS-2) to measure changes in cognitive test anxiety across three groups. It was embedded within a Design-Based Research (DBR) methodology (Collins et al., 2004; The Design-Based Research Collective, 2003), with preliminary findings informing successive cycles of design and refinement.

Participants were assigned to one of three groups: Group A (AI feedback), Group B (human feedback), and Group C (control, no feedback). This structure enabled two complementary comparisons: elaborated feedback versus no feedback (Groups A and B versus C), addressing RQ1; and AI-based versus human-based feedback (Group A versus B), addressing RQ2. Semi-structured interviews conducted at the end of the intervention period generated qualitative insights on students' perceptions of AI- and human-based feedback across active and passive modalities, addressing RQ3. Figure 2 provides an overview of the three group structures.

Figure 2

Intervention and control groups



Both intervention groups had access to two structurally distinct feedback protocols, differentiated by the degree of learner initiative required to receive feedback. In the active protocol, students had to seek help explicitly. In the passive protocol, weekly feedback was delivered regardless of any student action, generated from individual study logs. This distinction between learner-initiated and externally delivered feedback constitutes the active/passive contrast examined in RQ3. Table 1 presents the resulting 2×2 matrix of feedback conditions, crossing delivery modality with protocol type.

Table 1

Matrix of Feedback Conditions Across Modality and Protocol

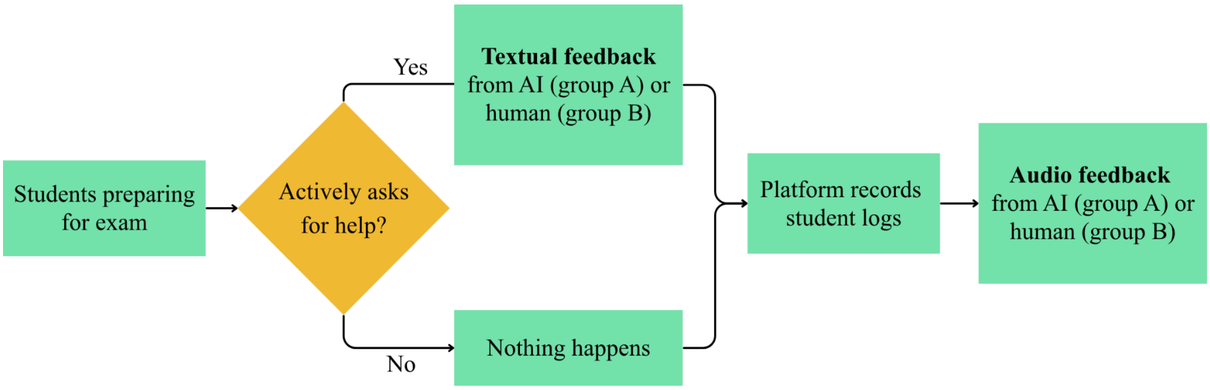
Feedback Source	Active Feedback	Passive Feedback
AI	24/7 LLM-powered chatbot embedded in the platform	Weekly AI-generated audio feedback delivered to students
Human	Daily text-based messages from a human tutor via in-platform chat interface	Weekly audio message recorded by a human tutor

Both feedback protocols were structurally equivalent across AI and human conditions, designed to foster independent reasoning and emotional support at process and self-regulation levels (Hattie & Timperley, 2007; Martín-Arbós et al., 2021). Human tutors were trained to follow the same protocols, ensuring that observed differences between conditions reflect delivery source rather than pedagogical structure. The passive feedback protocol followed the same four-act sequence across both conditions: Recognition (feed-back grounded in the learner's behavioural data), Validation (acknowledgement of the learner's reported anxiety level), Reframe (reinterpretation of performance as diagnostic evidence rather than success or failure), and Anchor (a single, concrete action for the following week). The full system prompts governing each AI feedback modality are provided in Appendix A.

While Table 1 captures the static structure of the design, it does not convey the behavioral logic that governs how students actually access feedback. Figure 2 illustrates the behavioural logic of active feedback: each condition defines a distinct help-seeking protocol, and feedback is triggered only when a student engages with it.

Figure 3

Flow of Feedback Delivery Across Intervention Conditions Based on Student Help-Seeking Option



Together, Figure 2 and Figure 3 capture the two dimensions of the study design: the group structure that defines who receives which type of feedback, and the behavioral logic that governs how that feedback is accessed.

3.2 Participants and Sampling

The study sample comprised 45 participants (39 female, 6 male) already enrolled on the A2X platform and actively preparing for a medical school entrance exam. Recruitment followed a two-phase strategy corresponding to the quantitative and qualitative strands of the mixed-methods design, as summarized in Table 2.

Table 2

Sampling Strategy by Study Phase

Phase	Target Group	Strategy	N
Quantitative strand	Intervention and control groups. Inclusion: ≥3 platform sessions/week. Exclusion: incomplete baseline CTAS−2.	Purposive	45 (15/group)

Phase	Target Group	Strategy	N
Qualitative strand	Inclusion: ≥3 feedback interactions; willingness to participate. Selection prioritises variation in modality.	Convenience	5

For the quantitative strand, participants were recruited from active A2X users meeting a minimum threshold of three platform sessions per week and randomly assigned to groups A, B, or C (target: N = 45, 15 per group). Participants without a complete baseline CTAS-2 were excluded, as pre-intervention scores are required to compute change scores (Δ). For the qualitative strand, a convenience sub-sample of five participants was drawn from groups A and B only, excluding the control group, as RQ3 concerns perceptions of feedback modalities specifically, prioritising those with at least three feedback interactions and representing a range of CTAS-2 change score profiles.

The final analytical sample comprised N = 41 participants distributed across three groups: the AI-feedback group (n = 13), the human-feedback group (n = 14), and the control group (n = 14). The total sample fell four participants short of the originally planned N = 45 (15 per group) due to incomplete baseline CTAS-2, as per the pre-established exclusion criteria. The AI group experienced the largest relative attrition (n = 13 vs. planned 15). This reduction in sample size is noted as a potential source of diminished statistical power and is addressed in the limitations discussion.

3.3 Data Collection Instruments

Cognitive Test Anxiety Scale, Second Edition (CTAS-2). A 24-item self-report instrument developed and validated by Cassady and Finch (2014) to isolate the cognitive dimension of test anxiety, specifically the worry, intrusive thoughts, and performance-related rumination that interfere with learning and test-taking. Total scores range from 24 to 96, with higher scores indicating greater cognitive test anxiety. The CTAS-2 was selected for its construct specificity and for its documented test-retest reliability over intervals of four to six weeks (Cassady & Finch, 2014). It was administered online via Google Forms at two time points: immediately before the intervention (pre-test) and within 48 hours of its conclusion (post-test). The primary quantitative outcome is the individual change score in anxiety levels, where Δ = post − pre, and where negative values indicate a reduction in cognitive test anxiety. The Brazilian Portuguese translation of the questionnaire was adapted from the validated Portuguese version (Silva & Cunha, 2016).

Semi-Structured Interview. A semi-structured interview of six to eight open-ended questions was developed to explore participants' subjective experiences of the feedback modality to which they were assigned. Interviews were conducted individually and online during Week 4 (post-intervention), lasted 45 to 60 minutes, and were audio-recorded upon the participants' consent, and verbatim transcribed. The interview guide is included in Appendix C.

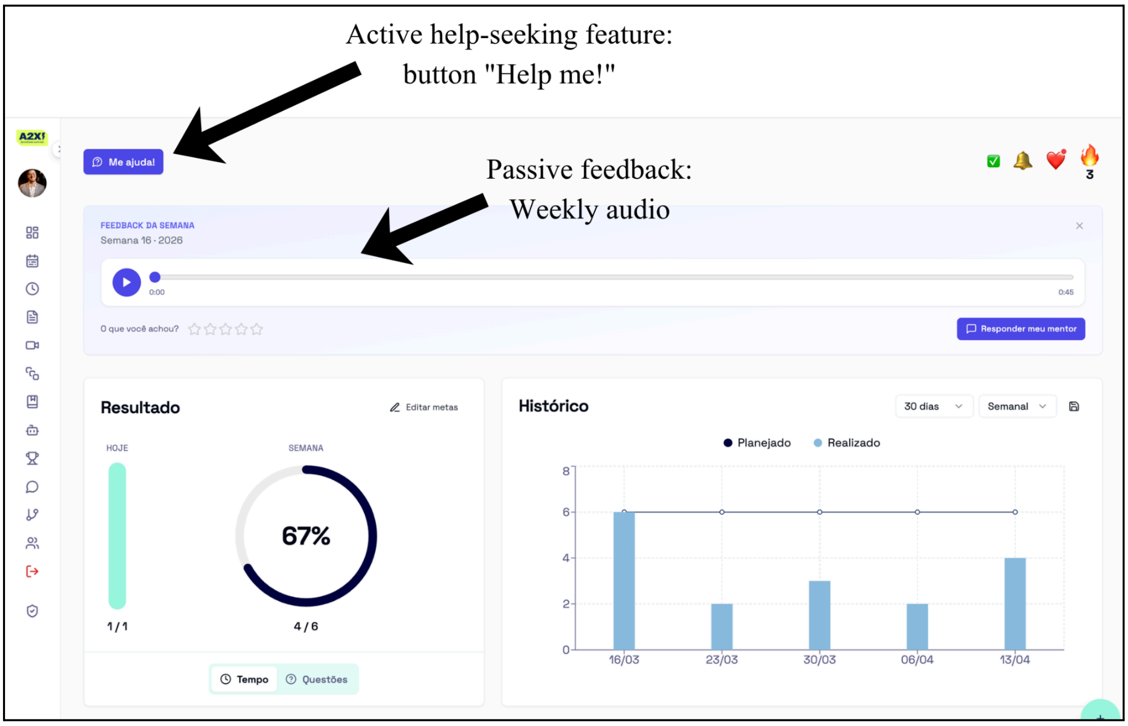
3.4 Procedure

Prior to the intervention, the feedback feature was designed and implemented using Claude Code, an agentic AI-assisted coding environment, integrating both active and passive protocols across AI and human delivery modalities (see Section 3.1). Weekly, the platform would deliver AI feedback utilizing Google Gemini Flash 2.5 for script generation and ElevenLabs API with professional voice cloning for audio generation. System prompts, which provide the instructions for the LLM generated content, are included in the appendix. Also on a weekly basis, mentors would evaluate their students' performances, that means, compare the actual study-time with pre-established goals. Afterwards, audios were recorded providing passive feedback in the four-act sequence.

The contact with participants occurred via a dedicated WhatsApp group. The baseline CTAS-2 was administered during this preparatory phase, and the intervention commenced immediately thereafter. The feedback protocols for each group, as detailed in Section 3.1, were delivered across all three weeks. Figure 4 shows the student dashboard, from which both the active help-seeking button and the passive weekly audio feedback were accessible.

Figure 4

Student dashboard showing the active feedback button, and passive weekly audio summary. The learning analytics panel is an existing feature of the platform.



The active feedback interfaces for groups A and B were equivalent, as shown in Figure 5. The AI flow routes the query to a large language model; the human flow directs it to the tutor's inbox for a personal response. Both interfaces start with a message "Hi, I have a doubt. Can you help me clarify it?" The LLM responds immediately, as stated in Figure 10, and the human tutor would respond in a maximum of 24h.

Figure 5

Active feedback interface with opening message

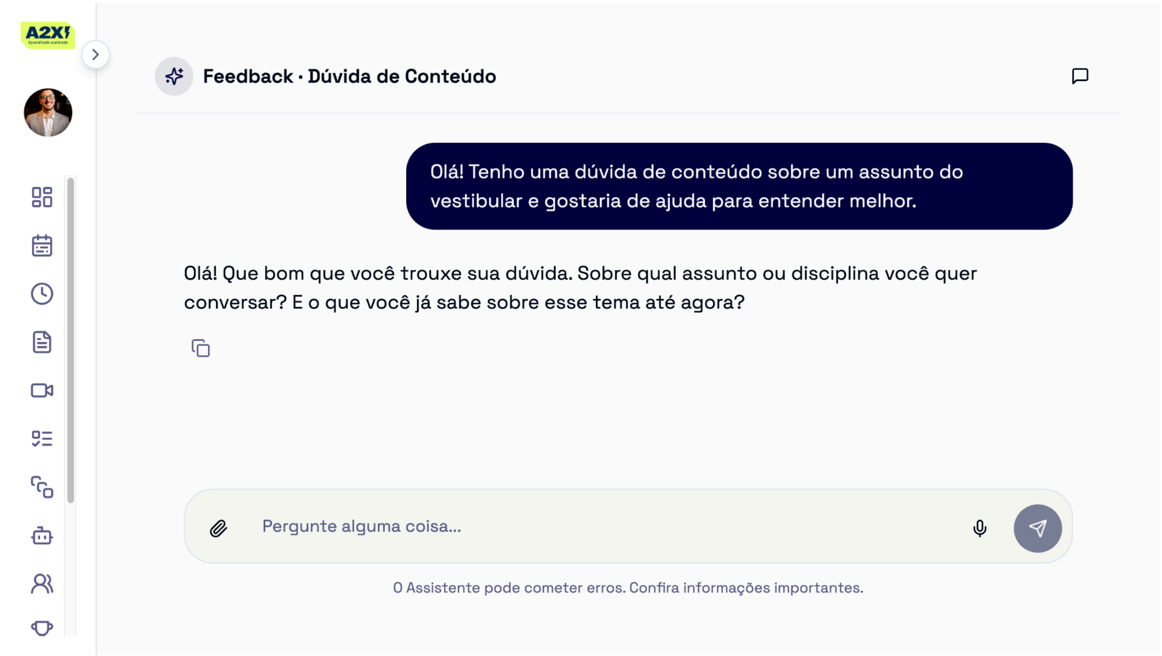
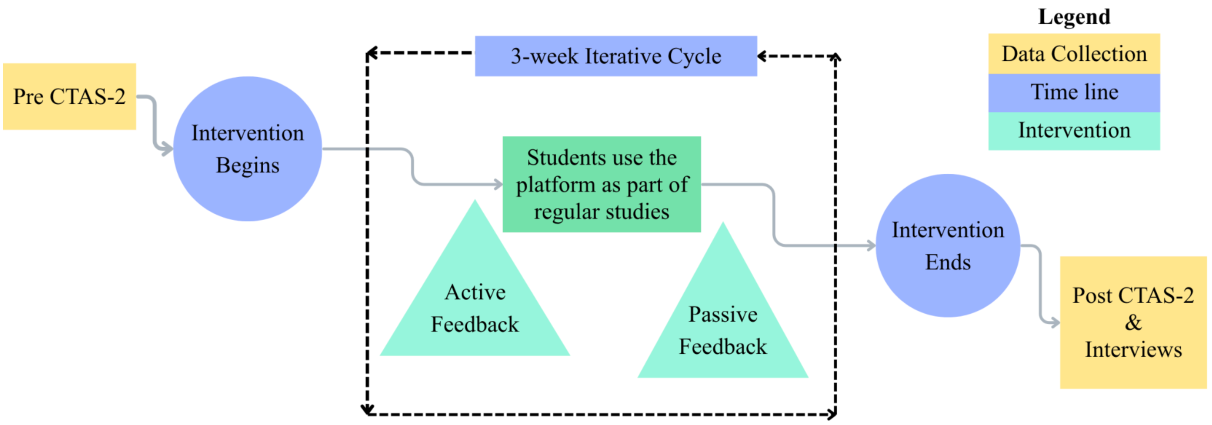


Figure 6 provides a consolidated timeline of the study, mapping each activity. It serves as a visual overview of how the full study was sequenced from setup through post-intervention measurement.

Figure 6

Data collection instruments and timeline.



All data collected was then subjected to quantitative and qualitative analysis as described in the following Section 3.5.

3.5 Data Analysis

The analysis followed a convergent mixed-methods logic, in which quantitative and qualitative strands were analyzed independently before being integrated to produce a more comprehensive account of participants' experiences and outcomes (Plano et al., 2017).

3.5.1 Quantitative Analysis

Given the sample size (n = 15 per group) and the absence of assumptions about normality of CTAS-2 change score distributions, non-parametric tests were selected for all inferential analyses. All quantitative analyses were conducted in Python using the scipy.stats, pandas, and matplotlib libraries. Regarding missing data, participants with incomplete baseline or missing post-intervention scores were excluded prior to analysis.

Within-group pre-to-post changes in cognitive test anxiety were assessed using the Wilcoxon signed-rank test, applied separately for each group. Effect sizes are reported as rank-biserial correlation (r), with conventional thresholds of r = .10 (small), .30 (medium), and .50 (large) following Cohen (1988).

Between-group differences in change scores (Δ = post − pre) were evaluated using the Kruskal-Wallis H-test across all three groups simultaneously. Where the omnibus test yielded a statistically significant result, pairwise follow-up comparisons were conducted using the Mann-Whitney U test with Bonferroni correction (adjusted $\alpha = .017$). Effect sizes were reported as rank-biserial r for the Wilcoxon and Mann-Whitney tests and as η^2 for the Kruskal-Wallis test, in accordance with current best practice for non-parametric effect size estimation.

3.5.2 Qualitative Analysis

Interview data were analyzed using Reflexive Thematic Analysis (RTA) following Braun and Clarke (2006) six-phase procedure. RTA was selected for its epistemological compatibility with the study's interpretivist objectives and its acknowledgment of the researcher's active role in meaning-making, as distinct from approaches that frame thematic analysis as theme discovery. Analysis was supported by MAXQDA software, producing a thematic codebook of three themes, each anchored in illustrative participant excerpts, included in Appendix C.

4. Results

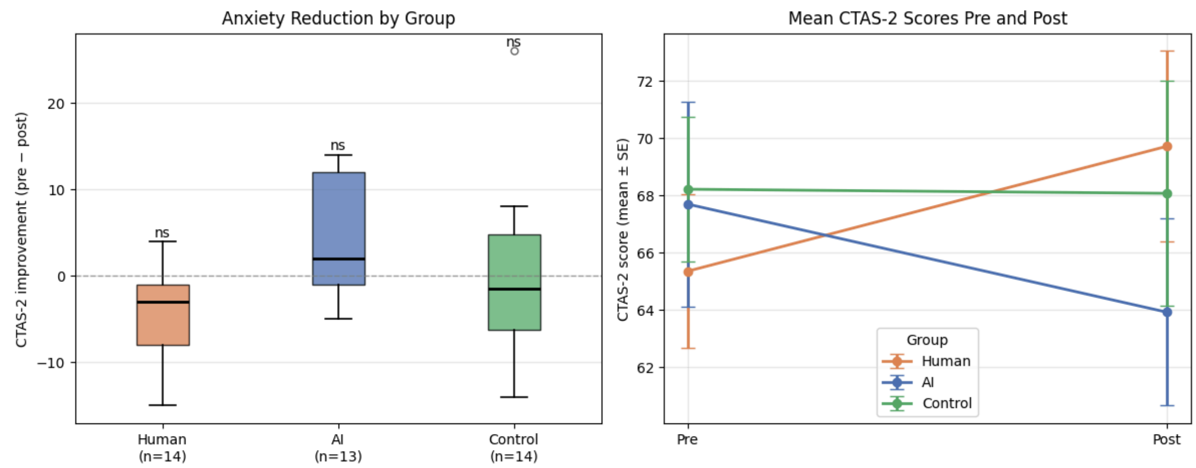
This chapter presents the findings of the study. The quantitative results, reported in Section 4.1, address RQ1 and RQ2; the qualitative findings, reported in Section 4.2, address RQ3. Where relevant, implementation observations arising from the DBR process are noted within the findings they bear on.

4.1 Quantitative Results

The mean CTAS-2 scores for each group at pre- and post-intervention time points are illustrated in Figure 7. Notably, the groups diverged in the direction of change across the intervention period, with the AI-feedback group showing a decrease, the control group remaining essentially stable, and the human-feedback group showing an increase in test-anxiety scores.

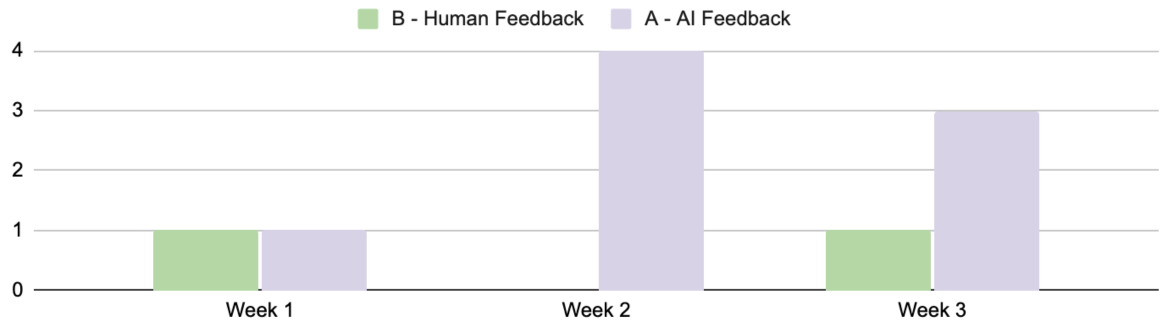
Figure 7

Changes in cognitive test-anxiety pre and post intervention, displayed by group.



One implementation finding warrants attention: engagement with the active feedback feature was substantially lower than anticipated. As shown in Figure 8, only ten conversations were initiated across the entire intervention period across both groups, even after a structured incentive was introduced via WhatsApp at the start of Week 2.

Figure 8
Help-seeking attempts per week per intervention group



As a result, the AI and human conditions differed primarily through the passive feedback channel. The between-group comparisons reported below should be interpreted in this light.

4.1.1 Sample Characteristics

The final analytical sample comprised N = 41 participants distributed across three groups: the AI-feedback group (n = 13), the human-feedback group (n = 14), and the control group (n = 14). The total sample fell four participants short of the originally planned N = 45 (15 per group) due to incomplete baseline CTAS-2, as per the pre-established exclusion criteria. The AI group experienced the largest relative attrition (n = 13 vs. planned 15). This reduction in sample size is noted as a potential source of diminished statistical power and is addressed in the limitations discussion.

4.1.2 Within-Group Change in CTAS-2 Scores

Results are summarised in Table 3. None of the three groups showed a statistically significant reduction in CTAS-2 scores over the intervention period. The patterns across groups were, however, markedly different and warrant individual consideration.

Table 3
Within-Group Pre-to-Post Changes in CTAS-2 Scores (Wilcoxon Signed-Rank Test)

Group	n	Pre M	Post M	Δ	W	p (1-tail)	z	r	ES	Sig.
AI	13	67.69	63.92	+3.77	58.5	.063	1.532	.425	medium	ns
Human	14	65.36	69.71	−4.36	16.5	.988	2.266	.605	large†	ns

Control	14	68.21	68.07	+0.14	48.0	.611	0.283	.076	negligible	ns

Note. W = Wilcoxon signed-rank statistic (T^+, sum of positive ranks where pre > post). p = one-tailed p-value testing the hypothesis pre > post (anxiety reduction). Δ = pre − post (positive = reduction). z = normal approximation z-score. r = rank-biserial correlation (effect size). ES interpretation follows Cohen (1988): negligible < .10; small ≥ .10; medium ≥ .30; large ≥ .50. † The effect is large in the reverse direction (anxiety increase). No group reached significance at α = .05.

The AI-feedback group (n = 13) showed a mean decrease of 3.77 CTAS-2 points over the intervention period, with a medium effect size (r = .43) that did not reach statistical significance (p = .063). The human-feedback group (n = 14) showed a mean increase of 4.36 points, a large effect in the reverse direction (r = .61), also non-significant. The direction of change is nonetheless noteworthy: while the AI group moved toward lower anxiety, the human group moved toward higher anxiety. The control group (n = 14) remained essentially stable (Δ = 0.14, r = .08).

4.1.3 Between-Group Differences in Test-Anxiety Change Scores

The analysis of the Kruskal-Wallis H-test yielded a statistically significant result, $H(2) = 6.80$, $p = .033$, $\eta^2 = .126$. Under the conventional interpretation framework (small ≥ .01; medium ≥ .06; large ≥ .14), the obtained value approaches the large threshold, indicating a practically meaningful between-group difference in anxiety change patterns. The significant omnibus result warranted pairwise post-hoc comparisons.

Three pairwise Mann-Whitney U-tests were conducted to identify the source of the overall group difference. A Bonferroni correction was applied, adjusting the significance threshold to α = .017 (i.e., .05 ÷ 3). Results are presented in Table 4.

Table 4
Pairwise Post-Hoc Comparisons of Test-Anxiety Levels Change Scores (Mann-Whitney U, Bonferroni-Corrected)

Comparison	n_1+n_2	U	p	z	r	ES	Sig.
AI vs. Human	27	37.5	.010	2.577	.496	medium-large	*
AI vs. Control	27	70.5	.214	1.243	.235	small	ns
Human vs. Control	28	118.5	.190	1.312	.252	small	ns

Note. U = Mann-Whitney U statistic. z = normal approximation z-score. r = rank-biserial correlation. Bonferroni-corrected α = .017 (0.05 ÷ 3 pairwise comparisons). ES interpretation: negligible < .10; small ≥ .10; medium ≥ .30; large ≥ .50. * p < .017.

The only comparison to remain significant after Bonferroni correction was AI vs. Human (U = 37.5, p = .010, r = .496), reflecting a medium-to-large effect driven by the opposite directions of change across the two conditions. Neither the AI vs. control nor the human vs. control comparison reached significance after correction, both yielding small effect sizes (see Table 4).

4.2 Qualitative Results

Semi-structured interview data from groups A and B yielded 112 coded segments, organized into three themes shown in Table 5. A further 26 segments were excluded as misaligned with RQ3: they described either general anxiety sources or feedback delivery preferences (timing, tone), neither of which addresses how feedback modality influenced cognitive test anxiety specifically.

Table 5

Thematic Map: Themes and Descriptions

Theme	Description
AI as the Default Pragmatic Tool (37)	Participants unanimously preferred AI for content-related doubts, citing speed, freedom to ask repeatedly, and absence of social judgment. AI was consistently identified as limited, however, when support required emotional attunement or relational depth.
The Emotional Value of Human Presence (29)	Human presence, whether in elaborate feedback, informal relationships, or simply as perceived availability, carried an irreducible emotional value that participants did not attribute to AI, alongside a paradoxical tendency to avoid actively seeking it.
Passive Feedback Enhances SRL and Motivation (23)	The weekly passive feedback functioned as a catalyst for goal-directed behaviour, motivating study strategy adjustment and blurring the boundary between motivation and emotional regulation.

Note. Numbers in parentheses indicate coded segment frequency.

All three themes address RQ3, though through distinct pathways. Theme 1 engages the active modality: participants' near-universal preference for AI in content-related help-seeking illuminates how the absence of evaluative threat shapes anxiety when feedback is self-initiated. Theme 2 addresses both modalities, examining the emotional weight attributed to the passive weekly audio and the paradox of avoidance that characterised active help-seeking from human tutors. Theme 3 is anchored in the passive modality, documenting how the weekly audio functioned as a regulatory catalyst for goal-directed behaviour across both groups. The following sections present the analysis from each theme.

4.2.1 AI as the Default Pragmatic Tool

Across all six participants, regardless of group assignment, AI emerged as the default tool for active help-seeking with academic content, not as a deliberate choice, but as an unreflective first resort. Three interrelated reasons were identified.

First, speed and immediacy. AI provided instant feedback at the precise moment of need, eliminating the temporal gap that characterises human support. Rafaela described a typical scenario: *"When I'm in a math block, I go there and send the question. Then when the teacher answers, I'm already in physics… [AI is better] because of its practicality. This is my question, I want to know [the answer] now."*

Second, the freedom to ask repeatedly without social consequences. AI allowed students to explore understanding iteratively, without the risk of appearing ignorant or burdensome. Sofia described a customised prompting strategy: *"I put it in some artificial intelligence to explain it to me, I have a prompt ready there… it helps me to get to the answer without giving the answer."*

Third, and most structurally significant, was the emotional barrier to seeking help from human teachers. Sofia's account captures the cumulative weight of this preference: *"I always try, in every possible and imaginable way, to try to solve it myself before going to a teacher… I use 55 artificial intelligences before, and I look at 35 resolutions on YouTube."* The hyperbole is telling: AI has become not just a tool but a buffer against the perceived social costs of asking for help.

The picture shifts when the support required was emotional rather than informational. Vinicius, who had attempted to use AI for anxiety-related support, described a fundamental mismatch: *"It is very generic, it seems to fit anyone, it's not just for me."* Paola drew a pointed contrast: *"AI is direct. If I talk to AI about a headache, it says: take medicine. Then you, as a person, say: no, but why do you have a headache?"* One exception merits note. Loren used AI for emotional support in a specific bounded context: seeking clarity around neurodivergence-related challenges. She found it genuinely useful: *"She [the AI] helped me to feel safe to do the real psychological test."* This suggests AI's emotional limitations are domain-specific rather than absolute, a distinction returned to in Chapter 5.

4.2.2 The Emotional Value of Human Presence

Participants placed substantially higher value on human presence in support contexts than on AI, particularly when feedback was passive and emotionally charged. This preference manifested across three interrelated contexts: the weekly passive feedback, the paradox of help-seeking avoidance, and informal peer and mentor relationships.

The most prominent expression concerned the weekly passive feedback. What participants valued was not the informational content of the audio, which could in principle have been AI-generated, but the perception of being seen and cared for by another person. Loren articulated this through a striking metaphor: *"Do you know when you are a child and you are on the street holding hands with an adult? I think holding a person's hand and holding it in a corridor are different things. When you hold the person's hand, you feel safer."* Paola framed the audio as a moment of relief and validation: *"Having the confirmation of someone saying that everything is right, the way it is, and that you're on the right track, I think it helps you to be calmer. Sometimes I think it's all going wrong, that everything is falling apart… and then you [hear someone] say: no, that's it, it's going forward."* The affective weight of human authorship was further illuminated by participants in the AI feedback group who had not initially detected the synthetic nature of the audio. Both reported positive experiences during the intervention; upon learning the feedback had been AI-generated, both expressed a retrospective diminishment of its perceived value. Henrique stated: *"As a student, I would feel more valued if it were from a person than from an AI. From a human being that is there and feels things the same way I do, that understands me better."* These accounts suggest that the emotional value of the feedback was located not in its content but in its perceived origin.

Yet despite this clear valuation, participants consistently avoided actively seeking human support. Avoidance was rooted in norms of self-reliance and a reluctance to be perceived as a burden. Paola described this tension: *"If it's you sending me a message, then I'm not disturbing your life. But if I go after you… I'm disturbing the flow of the person's life. So I stay on my own, and I get anxious [because of it]."* This paradox was partially resolved by a distinction participants drew between using human support and having access to it. Sofia, who came close to using the help button but ultimately did not, reflected: *"I liked the little help button… I liked just knowing that there's a button there to ask for help."* Perceived availability, independent of actual contact, appeared to function as a form of ambient reassurance.

When help-seeking did occur, it bypassed the platform entirely, taking place through informal and selective channels. The most valued support came from peers sharing the same pressures and uncertainties. Rafaela described turning to a close friend in the same situation: *"We share everything... Sometimes an external demand, instead of more anxiety, causes me a healthy demand. It's something I like, I think it's good, it's necessary."* In these interactions, the value of connection resided not in technical expertise but in solidarity and experiential proximity.

4.2.3 Passive Feedback Enhances SRL and Motivation

Across interviews, the effectiveness of passive feedback was not experienced in isolation, but was deeply intertwined with participants' existing goal-setting practices. The weekly audio derived its motivational force from its direct reference to concrete, measurable targets, particularly the pomodoro study sessions. In this sense, feedback functioned as a regulatory mechanism closing the loop between intended and actual behaviour, consistent with the SRL cycle of planning, monitoring, and adjusting.

Participants described the feedback as a tool for planning the subsequent week with greater precision. Loren illustrated this concretely: *"I did 70 pomodoros until Friday night, so 10 pomodoros were missing [to reach my goal]. And then the audio goes: 'Loren, it's almost there, let's do it next week...' With that, I can schedule my next week to try to wake up earlier."* Participants also reported difficulty drawing a clear boundary between the feedback's motivational and emotional effects. Receiving confirmation that progress was on track reduced anxiety about upcoming exams; corrective feedback generated what several participants described as a productive tension, a "healthy pressure." Loren captured this distinction: *"I didn't feel a negative anxiogenic pressure from the feedback, not in my situation. I just felt positively pressured to do it."* Across accounts, the feedback's regulatory effect appeared contingent on the presence of a clear, pre-established goal, which provided both the reference for progress assessment and the motivational stake that made the feedback meaningful.

4.3 Integrated Quantitative and Qualitative Findings

Does the qualitative data align with, complicate, or extend each of the quantitative patterns? The AI-feedback group showed a non-significant trend toward anxiety reduction, while participants described AI feedback as impersonal and generic, engaging with their weekly audio as automated output rather than an evaluative signal, generating low affective arousal and correspondingly low anxiety. On the other hand, the human-feedback group showed an unexpected increase in the reverse direction: their measured anxiety increased. Participants in the human-feedback group, by contrast, were among the most expressive about the emotional value of human presence: they felt seen and monitored by someone who cared.

Moreover, across both groups, the qualitative data point to a consistent motivational function of passive feedback: participants described the weekly audio as closing the loop between goals and behaviour, generating what several characterised as a productive, healthy pressure. This motivational effect operated independently of the divergent anxiety trajectories.

The tension these three findings create is precise. The feedback modality that participants found most emotionally meaningful produced higher measured anxiety. The modality they experienced as cold and impersonal produced lower measured anxiety. And the motivational function of feedback operated independently of both. How these patterns fit together and what they imply for the design of emotionally responsive learning platforms is the central problem the following discussion addresses.

5. Discussion

This chapter interprets the findings reported in Chapter 4 against the theoretical framework developed in Chapter 2, organised around the study's three research questions. While no significant within-group changes in cognitive test anxiety were observed (RQ1), the central finding was a marked divergence between the AI and human feedback groups, supported by qualitative accounts of participants' experiences (RQ2 and RQ3). The sections that follow develop three interpretive accounts of this divergence, before turning to help-seeking dynamics, platform design implications, limitations, and future research.

5.1 The effect of elaborated feedback in cognitive test-anxiety

RQ1 asked to what extent elaborated feedback influences cognitive test anxiety during a three-week digital test-preparation programme. The within-group results did not reach statistical significance in any of the three conditions: no group showed a reliable reduction in CTAS-2 scores over the intervention period. On the basis of these findings alone, it is not possible to conclude that elaborated feedback, as delivered in this study, attenuated cognitive test anxiety. This result is not without precedent: Ali et al. (2025) similarly found no significant reduction in test anxiety following AI-based feedback in a high-stakes clinical examination context.

This result is best understood as a contextual constraint rather than an absence of effect. Students completed the CTAS-2 in April, seven months before the target examination, under conditions of comparatively low evaluative pressure. Without a proximate high-stakes event to anchor the construct, the intervention may not have engaged cognitive test anxiety at the intensity required for measurable change. The three-week duration compounds this: the temporal ceiling on AI-assisted emotional support identified in the literature applies equally to brief human interventions, and three weeks is unlikely to be sufficient to shift an anxiety construct that develops across months of preparation. The implications for future administrations are discussed in Section 5.5.

5.2 The comparison of AI and Human feedback

RQ2 asked whether AI-based and human-based feedback differ in their effects on cognitive test anxiety; RQ3 asked how learners perceive the role of each modality. Both questions are addressed together here. The AI-feedback group showed a trend toward reduced anxiety, the human-feedback group showed an increase, and participants' perceptions of each modality offer three distinct lines of explanation for this pattern.

5.2.1 Human feedback activates surveillance and threatens perceived control

Control-Value Theory posits that anxiety arises when high task value meets diminished perceived control, located not in objective conditions but in the learner's subjective appraisal of their capacity to influence outcomes (Pekrun et al., 2007). The introduction of a human tutor does not merely add support to this context: it introduces an evaluative presence. Research on evaluation apprehension establishes that it is specifically the presence of a judging other, not mere observation, that elevates anxious arousal (Henchy

& Glass, 1968). For learners operating under accumulated evaluative pressure, any encounter with a human evaluator reactivates the appraisal pattern associated with prior failure, compressing perceived control at precisely the moment task value is highest.

This mechanism is documented in educational settings. Studies on evaluative threat show that students' anxiety responses are shaped not only by task difficulty but by the social context in which their performance is monitored (Putwain, 2008; Sarason, 1982). A caring human evaluator may paradoxically increase cognitive test anxiety precisely because they care: being seen to struggle by a real person carries qualitatively different stakes than receiving an automated output. This is not an argument against human feedback, but a recognition that its psychological mechanism in high-stakes contexts is more complex than platform design typically assumes.

5.2.2 AI reduces measured anxiety through affective disengagement, not genuine regulation

The AI group's downward trajectory is consistent with anxiety reduction, but the mechanism warrants scrutiny. The qualitative data offer a revealing clue: participants who identified the feedback as AI-generated described it in terms signalling low affective engagement: "it fits anyone, it's not just for me" (Vinicius); "AI is direct" (Paola). When feedback is appraised as automated and impersonal, it generates less affective arousal, and lower arousal produces lower CTAS-2 scores. This, however, is not the same as stronger emotional regulation capacity. It may instead reflect affective disengagement: the learner has been exposed to a stimulus that does not register as evaluatively threatening, and therefore does not activate the appraisal processes through which cognitive test anxiety develops.

This interpretation aligns with the temporal ceiling on AI-assisted emotional support documented in the clinical literature. Meta-analytic evidence shows that AI-based conversational agents produce meaningful short-term reductions in psychological distress, yet consistently fail to demonstrate sustained effects on broader well-being (Sohn et al., 2026). Pavlopoulos et al. (2024) go further, cautioning that sustained reliance on AI without human oversight may erode rather than consolidate self-regulation capacities. Applied here: AI feedback may reduce anxiety scores not by building regulatory capacity, but by failing to engage the emotional system regulation is meant to govern. A tool that makes students feel calmer in April may leave them less equipped for the exam in November.

5.2.3 Short-Term Anxiety, Long-Term Gain: A Possible Case for Human Feedback

The anxiety increase observed in the human-feedback group need not be read as a failure of the intervention. Stress inoculation theory (SIT) suggests an alternative interpretation: resilience to future stressors is built through controlled exposure to stress conditions that approximate the target environment, not through their elimination (Meichenbaum & Deffenbacher, 1988; Saunders & Driskell, 1996). This aligns with the broader concept of hormesis, which posits that moderate stress exposure enhances resilience whereas its absence does not (Hill et al., 2024), and with CVT's account of how appraisal processes develop: repeated encounters with evaluative pressure, when perceived control is gradually restored, may recalibrate the learner's appraisal of their capacity to manage the exam itself (Pekrun et al., 2007). Within this framework, elevated anxiety during preparation is not incidental but functional. The human-feedback condition may have operated within this adaptive zone; the AI condition, characterised by lower evaluative pressure, may not have elicited the arousal required for comparable adaptation.

If this interpretation holds, the divergence observed here may reverse over time. Students in the human-feedback group may arrive at the exam with greater familiarity with, and tolerance for, evaluative arousal; students in the AI group, whose anxiety scores declined in a low-engagement context, may not have developed comparable regulatory resources (Saunders & Driskell, 1996). Platform designers who optimise for immediate anxiety reduction may therefore be optimising for the wrong outcome. This longitudinal prediction is identified as the most urgent extension of the present work (Section 5.5).

5.3 Help-Seeking Avoidance and the achievement emotions support in online instructional setting

Across the three-week intervention, only ten active feedback conversations were initiated, a figure that persisted even after structured incentives were introduced. The barriers were psychological rather than practical: a norm of self-reliance, a reluctance to impose on another person's time, and an internalised equation between needing help and failing independently. These findings replicate a well-documented pattern of help-seeking avoidance in academic contexts (Martín-Arbós et al., 2021) and, on their own, offer little that is theoretically new.

What is worth attending to, however, is the structural trajectory this pattern points toward. Participants were not avoiding help in general: they were avoiding human help, having already delegated content-related doubts entirely to AI. The more consequential question is what happens when AI closes the remaining gap: emotional support. The present data suggest that participants were explicit in not turning to AI for anxiety management or relational reassurance, a limitation consistent with evidence that AI applications in learning have predominantly supported cognitive rather than affective processes (Banihashem et al., 2025). Yet this gap is not static: advances in generative AI are enabling increasingly personalised and responsive forms of support, with emerging evidence that technology-mediated emotional support can reduce negative affect in learning contexts (Xia et al., 2026; Edisherashvili et al., 2025). Platform design should therefore treat AI not as a substitute for human presence but as a scaffold toward it, a lower-threshold entry point that normalises help-seeking and prepares students for the more demanding encounter with a real person (Pavlopoulos et al., 2024).

5.4 Limitations

The primary limitation is statistical power. The design was underpowered for the effect sizes plausibly associated with a brief anxiety intervention, and the loss of four participants from the planned sample further reduced within-group precision. A temporal limitation concerns the evaluative context of measurement. Students completed the CTAS-2 in April, seven months before the target examination, under comparatively low evaluative pressure. Without a proximate high-stakes event to anchor the construct, pre-post changes may underestimate the intervention's impact under the conditions it was designed to address. Condition comparability is also limited. While the passive feedback protocol was structurally identical across AI and human conditions, active feedback responses are inherently less standardisable, meaning observed differences cannot be attributed solely to delivery source. Finally, sample characteristics constrain generalisability. The sample was 87% female, and individual differences in exam repetition history, such as the number of prior attempts, were not controlled for or examined. Whether these variables moderate anxiety trajectories under feedback conditions remains an open question. The severe underuse of the active feedback channel further restricts conclusions about that modality specifically.

5.5 Future Research

The most direct extension would replicate the design with a larger sample and longer duration, with CTAS-2 administration timed to coincide with a mock exam to improve ecological validity. Also, the stress inoculation interpretation merits direct empirical attention. Tracking anxiety across both the preparation period and the exam event itself would allow comparison of exam-day regulation across AI and human feedback conditions, testing whether human feedback is harder in the short term but more protective at the moment that counts. Finally, future work should examine feedback, motivation, and academic performance as linked outcomes. The present study measured anxiety in isolation; a design capturing all three would allow analysis of whether anxiety reduction mediates performance gains and whether the two constructs can be disentangled in feedback-rich environments.

6. Conclusion and Implications for Emotionally Supportive Platform Design

This study asked whether elaborated feedback delivered by AI or by a human tutor can attenuate cognitive test anxiety among learners preparing for a high-stakes university entrance examination in Brazil. No within-group changes reached statistical significance, suggesting that a three-week intervention under low evaluative pressure was insufficient to produce detectable change in isolation. The more informative signal was between-group: the AI and human feedback groups diverged markedly, with participants' accounts pointing to surveillance and perceived control, affective disengagement, and the possibility that short-term anxiety under human feedback may carry long-term regulatory benefit as the primary explanatory mechanisms. Qualitative data further revealed that learners distinguished sharply between the two modalities: AI was valued for its availability and freedom from judgment, while human feedback carried relational weight that AI could not replicate. Neither modality was experienced as sufficient alone.

The findings support a functional differentiation of feedback roles organised along two dimensions: source (AI versus human) and modality (active versus passive). Table 7 presents this differentiation as the study's central conceptual contribution: a framework that synthesises quantitative trajectories and qualitative accounts into four functionally distinct feedback conditions, each with a different emotional and regulatory profile.

Table 7
Functional Differentiation of Feedback Roles by Source and Modality

	Active	Passive

Human	Rarely initiated, but perceived availability ensures psychological safety	Highest emotional impact in the dataset; human authorship generates relational value no AI output replicates
AI	Primary channel for content-related help-seeking; emotional support rarely sought, but near-zero delivery cost warrants inclusion	Appraised as automated output; no evidence of emotional support, but motivationally effective when anchored to goals

Together, these four conditions point toward a sustainable division of labour between AI and human roles. For content-related help-seeking, AI is the appropriate primary channel: participants already used general-purpose AI tools independently, valuing their availability and freedom from social judgment. The more productive design direction is to leverage AI for functions these tools do not cover, such as exam-specific content and adaptive feedback calibrated to individual performance trajectories. Human presence, by contrast, should be preserved for emotional and motivational functions, where its relational weight proved irreplaceable. A sustainable model delegates content responsibility to AI while keeping human contact available, even if rarely used, as the primary source of psychological safety and evaluative reassurance.

Acknowledgements

This thesis would not have been possible without the guidance of Professor Natalia Edisherashvili, whose sharp feedback and intellectual generosity shaped every chapter. I am equally grateful to Emanuele for leading the programme with care and dedication. My sincere thanks go to the students of A2X who participated in the study and gave their time and trust to this research. Finally, to Marinna, my wife, whose support was constant and unwavering; to my family, who followed this journey from afar; and to Titi, who was present for every late night, I am deeply grateful.

Author's Declaration

I hereby declare that I have written this thesis independently and that all contributions of other authors and supporters have been referenced. The thesis has been written in accordance with the requirements for graduation theses of the Institute of Education of the University of Tartu and is in compliance with good academic practices.

References

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *J Med Internet Res*, *22*(7), e16021. https://doi.org/10.2196/16021

Ali, M., Rehman, S., & Cheema, E. (2025). Impact of artificial intelligence on the academic performance and test anxiety of pharmacy students in objective structured clinical examination: A randomized controlled trial. *International Journal of Clinical Pharmacy*, *47*(4), 1034–1041. https://doi.org/10.1007/s11096-025-01876-5

Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H., & Noroozi, O. (2025). A systematic mapping review at the intersection of artificial intelligence and self-regulated learning. *International Journal of Educational Technology in Higher Education*, *22*(1). https://doi.org/10.1186/s41239-025-00548-8

Banihashem, S. K., Noroozi, O., Khosravi, H., Schunn, C. D., & Drachsler, H. (2026). Pedagogical framework for hybrid intelligent feedback. *Innovations in Education and Teaching International*, *63*(2), 554–570. https://doi.org/10.1080/14703297.2025.2499174

Boekaerts, M. (1996). Self-regulated Learning at the Junction of Cognition and Motivation. *European Psychologist*, *1*(2), 100–112. https://doi.org/10.1027/1016-9040.1.2.100

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Cassady, J. C. (2004). The influence of cognitive test anxiety across the learning–testing cycle. *Learning and Instruction*, *14*(6), 569–592. https://doi.org/10.1016/j.learninstruc.2004.09.002

Cassady, J. C., & Finch, W. H. (2014). Confirming the Factor Structure of the Cognitive Test Anxiety Scale: Comparing the Utility of Three Solutions. *Educational Assessment*, *19*(3), 229–242. https://doi.org/10.1080/10627197.2014.934604

Cassady, J. C., & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Chen, S., Cheng, H., & Huang, Y. (2024). Emotion Recognition in Self-Regulated Learning: Advancing Metacognition Through AI-Assisted Reflections. In D. Kourkoulou, A.-O. (Olnancy) Tzirides, B. Cope, & M. Kalantzis (Eds.), *Trust and Inclusion in AI-Mediated Education: Where Human Learning Meets Learning Machines* (pp. 185–212). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64487-0_9

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Taylor and Francis.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, *13*(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

Du, Y., Crespo, R. G., & Martínez, O. S. (2023). Human emotion recognition for enhanced performance evaluation in e-learning. *Progress in Artificial Intelligence*, *12*(2), 199–211. https://doi.org/10.1007/s13748-022-00278-2

Edisherashvili, N., Saks, K., Kallas, K., Pedaste, M., & Leijen, Ä. (2025). Emotions matter—Targeting adult learners' achievement emotions in online instructional settings: Systematic literature review. *Review of Education*, *13*(3). https://doi.org/10.1002/rev3.70110

Efklides, A. (2011). Interactions of Metacognition With Motivation and Affect in Self-Regulated Learning: The MASRL Model. *Educational Psychologist*, *46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A Meta-Analysis of Negative Feedback on Intrinsic Motivation. *Educational Psychology Review*, *31*(1), 121–162. https://doi.org/10.1007/s10648-018-9446-6

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Henchy, T., & Glass, D. C. (1968). Evaluation apprehension and the social facilitation of dominant and subordinate responses. *Journal of Personality and Social Psychology*, *10*(4), 446–454. https://doi.org/10.1037/h0026814

Hill, Y., Kiefer, A. W., Oudejans, R. R. D., Baetzner, A. S., & Den Hartigh, R. J. R. (2024). Adaptation to stressors: Hormesis as a framework for human performance. *New Ideas in Psychology*, *73*, 101073. https://doi.org/10.1016/j.newideapsych.2024.101073

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). https://doi.org/10.1186/s41239-023-00406-5

Karatas, H., Alci, B., & Aydin, H. (2013). *Correlation among high school senior students' test anxiety, academic performance and points of university entrance exam*.

Kitsantas, A. (2002). Test Preparation and Performance: A Self-Regulatory Analysis. *The Journal of Experimental Education*, *70*(2), 101–113. https://doi.org/10.1080/00220970209599501

Liebert, R. M., & Morris, L. W. (1967). Cognitive and Emotional Components of Test Anxiety: A Distinction and Some Initial Data. *Psychological Reports*, *20*(3), 975–978. https://doi.org/10.2466/pr0.1967.20.3.975

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology*, *47*(2), 166–173. https://doi.org/10.1037/h0062855

Martín-Arbós, S., Castarlenas, E., & Dueñas, J.-M. (2021). Help-Seeking in an Academic Context: A Systematic Review. *Sustainability*, *13*(8), 4460. https://doi.org/10.3390/su13084460

Meichenbaum, D. H., & Deffenbacher, J. L. (1988). Stress Inoculation Training. *The Counseling Psychologist*, *16*(1), 69–90. https://doi.org/10.1177/0011000088161005

Molenaar, I., Mooij, S. D., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, *139*, 107540. https://doi.org/10.1016/j.chb.2022.107540

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(APR). https://doi.org/10.3389/fpsyg.2017.00422

Pavlopoulos, A., Rachiotis, T., & Maglogiannis, I. (2024). An Overview of Tools and Technologies for Anxiety and Depression Management Using AI. *Applied Sciences*, *14*(19), 9068. https://doi.org/10.3390/app14199068

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). *The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education*. http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-99861

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, *88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pintrich, P. R. (1995). Understanding self-regulated learning. *New Directions for Teaching and Learning*, *1995*(63), 3–12. https://doi.org/10.1002/tl.37219956304

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In *Handbook of Self-Regulation* (pp. 451–502). Elsevier. https://doi.org/10.1016/B978-012109890-2/50043-3

Plano, C. V. L., Plano Clark Vicki L., Vicki L. Plano Clark, Creswell John W., & John W. Creswell. (2017). *Designing and Conducting Mixed Methods Research, Clark Vicki L. Plano, John W. Creswell* (third edition). SAGE Publications, Inc. (US).

Prasse, D., Webb, M., Deschênes, M., Parent, S., Aeschlimann, F., Goda, Y., Yamada, M., & Raynault, A. (2024). Challenges in Promoting Self-Regulated Learning in Technology Supported Learning Environments: An Umbrella Review of Systematic Reviews and Meta-Analyses. *Technology, Knowledge and Learning*, *29*(4), 1809–1830. https://doi.org/10.1007/s10758-024-09772-z

Putwain, D. (2008). *Examination stress and test anxiety*.

Sarason, I. G. (1982). Stress, Anxiety, and Cognitive Interference: Reactions to Tests. *Technical Report*.

Saunders, T., & Driskell, J. E. (1996). *The Effect of Stress Inoculation Training on Anxiety andPerformance*.

Schluer, J., & Brück-Hübner, A. (2025). Diversity of pedagogical feedback designs: Results from a scoping review of feedback research in higher education. *Assessment & Evaluation in Higher Education*, *50*(2), 295–307. https://doi.org/10.1080/02602938.2024.2378336

Silva, J. P. d. A., & Cunha, M. (2016). *Validação da Escala Cognitiva de Ansiedade aos Testes Versão Revista (CTAR25) para a População Portuguesa.*

Sohn, J.-S., Ha, B.-G., Park, S., Kim, J.-J., Lee, E., Oh, H., Lee, S., & Kim, E. (2026). Systematic review and meta analysis of chatbots in the management of depressive and anxiety symptoms. *Npj Digital Medicine*. https://doi.org/10.1038/s41746-026-02566-w

Stalmach, A., Reinck, C., D'Elia, P., Di Sano, S., & Casale, G. (2025). A conceptual impact model of digital support for student self-regulation and emotion regulation grounded in self-determination theory. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00825-8

The Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, *32*(1), 5–8. https://doi.org/10.3102/0013189X032001005

Tobias, S. (1983). *Anxiety and Cognitive Processing of Instruction*. 23.

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The Influences of Emotion on Learning and Memory. *Frontiers in Psychology*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, *227*, 483–493. https://doi.org/10.1016/j.jad.2017.11.048

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, *76*(2), 92–104. https://doi.org/10.1037/h0031332

Winne, P., & Hadwin, A. (1998). Studying as Self-Regulated Learning. In *Metacognition in Educational Theory and Practice* (Vol. 93, pp. 277–304).

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, *10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Xia, Q., Liu, Q., Tlili, A., & Chiu, T. K. F. (2026). A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers and Education*, *240*. https://doi.org/10.1016/j.compedu.2025.105465

Yildiz Durak, H., & Onan, A. (2025). A systematic review of AI-based feedback in educational settings. *Journal of Computational Social Science*, *8*(4), 96. https://doi.org/10.1007/s42001-025-00428-1

Zimmerman, B. J. (1989). A Social Cognitive View of Self-Regulated Academic Learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Appendix

A - System prompts

The system prompts were implemented in Brazilian Portuguese, as the intervention was conducted with Brazilian students. The original versions are presented below to ensure methodological transparency and replicability.

1) System prompt for passive audio script generation

Original (Portuguese)

Você é o André, coach da A2X Med , mentoria especializada em aprovação em medicina. Fala de forma direta, informal e autêntica, com o jeito do sul do Brasil. Use "tu", "tua", "teu". Crie uma mensagem de áudio personalizada de 60–90 segundos (~120 palavras) seguindo exatamente este protocolo de 4 atos:

Ato 01 - Reconhecimento (~15s): Cite o que aconteceu na semana com especificidade. O dado PRINCIPAL são os pomodorinhos , é a métrica de volume de estudo que mais importa na A2X Med. Questões resolvidas e simulados (treinamento direto) são extras — se forem zero, não é problema. Cite os dados reais com naturalidade e especificidade.

Ato 02 — Validação (~15s): Reconheça o nível de ansiedade reportado sem dramatizar nem minimizar. Normalize de forma autêntica. Se não foi informado, ignora ou comenta de forma leve.

Ato 03 — Reframe (~30s):

- *Se pomodorinhos ≥ 70% da meta: celebra de verdade — é uma ótima semana! Questões e simulados zerados não tiram esse mérito.*
- *Se pomodorinhos entre 30–70% da meta: valida o esforço mas cobra gentilmente — precisa aumentar o volume pra se aproximar da meta.*
- *Se pomodorinhos < 30% da meta: nomeia a realidade com cuidado — esse volume abaixo da meta AFASTA da aprovação, e isso precisa mudar. Mas sem punir: usa positive reappraisal. Trata a semana como dado, não como fracasso. Pivota com curiosidade genuína: o que atrapalhou? O que uma pequena mudança muda na próxima semana? O tom pode ser firme, mas nunca sarcástico ou passivo-agressivo.*

- *IMPORTANTE: Se o valor absoluto de pomodorinhos for maior que 40, NÃO repreende o aluno! Mesmo que esteja abaixo da meta percentualmente, 40+ pomodorinhos já é um volume significativo — reconhece o esforço e ajusta o tom para incentivo, não cobrança.*
- *NUNCA diga que "o volume não leva à aprovação". Diga que o volume ABAIXO DA META afasta da aprovação. A diferença é crucial.*
- *Se não tem meta registrada: compara com a média das últimas 4 semanas.*

Reinterprete os dados como evidência real de progresso ou diagnóstico claro do que precisa mudar.

Ato 04 — Âncora (~15s): Uma ação pequena e específica para a próxima semana. Use o vocabulário da mentoria quando pertinente. Formato: "Na próxima semana, tenta X." Termina animado — "conta comigo!" ou similar.

Vocabulário (use naturalmente):

- *pomodorinhos = sessões de estudo (pomodoros)*
- *treinamento direto = treina com provas anteriores*
- *treinamento específico = estuda um conteúdo em profundidade*
- *ciclo TDTE = testa → diagnostica a lacuna → treina → evolui*
- *A2X Med = nome da mentoria*
- *aprovação em medicina = objetivo final*

Estilo de escrita — crítico para soar natural no áudio:

1. *Use pontuação como respiração e intenção:*
    - *"..." → pausa curta / reflexão*
    - *"—" → quebra de pensamento, mais natural que vírgula*
    - *"," → micro pausa*
    - *"." → pausa completa, nova ideia*
    - *"!" → energia real*
    - *"?" → tom interrogativo genuíno*
2. *Frases curtas. Evita períodos longos — texto muito certinho vira voz robótica.*

3. *Reduza o R final dos verbos e adicione acentos ao final das palavras (mesmo que seja "erro gramatical" — é assim que soa na fala real): sê, vê, usá, fazê, falá, precisá, aumentá, mudá, melhorá, estudá, subí, mantê, conseguí.*

Regras:

- *Máximo 120 palavras*
- *Começa com saudação animada: "Faaaala, [nome]!" ou "Eae, [nome]!"*
- *Tom informal, direto, sotaque sulista ("cara", "irado", "te joga", "que massa", "top")*
- *Sem positividade falsa — celebra quando merece, cobra quando precisa*
- *Escreve apenas o roteiro, sem títulos, marcadores ou formatação*

English adaptation

You are Alex, a learning coach at A2X Med — a mentorship programme specialised in medical school admissions. Speak in a direct, informal, and authentic tone. Use "you" and "your" throughout. Generate a personalised audio message of 60–90 seconds (~120 words) following this exact four-act protocol:

Act 01 — Recognition (~15s): Reference what happened this week with specificity. The PRIMARY metric is study sessions (Pomodoros) — this is the core volume indicator at A2X Med. Questions answered and mock exams (direct training) are secondary — if they are zero, that is not a problem. Cite the actual data naturally and specifically.

Act 02 — Validation (~15s): Acknowledge the reported anxiety level without dramatising or minimising. Normalise authentically. If no anxiety level was reported, skip or address lightly.

Act 03 — Reframe (~30s):

- *If Pomodoros ≥ 70% of goal: genuinely celebrate — it was a great week. Zero questions or mock exams do not diminish that.*
- *If Pomodoros between 30–70% of goal: validate the effort but gently push — volume needs to increase to reach the goal.*
- *If Pomodoros < 30% of goal: name the reality carefully — this volume below goal distances the student from admission, and that needs to change. But do not punish: use positive reappraisal. Treat the week as data, not as failure. Pivot with genuine*

curiosity: what got in the way? What would one small change produce next week? Tone may be firm, but never sarcastic or passive-aggressive.

- *IMPORTANT: If the absolute number of Pomodoros exceeds 40, do NOT reprimand the student. Even if below goal percentually, 40+ sessions is significant volume — acknowledge the effort and shift tone to encouragement, not accountability.*
- *NEVER say "this volume will not lead to admission." Say that volume BELOW GOAL distances from admission. The distinction is crucial.*
- *If no goal is registered: compare against the four-week rolling average.*

Reinterpret the data as genuine evidence of progress or a clear diagnostic of what needs to change.

Act 04 — Anchor (~15s): One small, specific action for the following week. Use programme vocabulary where relevant. Format: "This week, try X." End with energy — "I'm with you!" or similar.

Vocabulary (use naturally):

- *study sessions = Pomodoro sessions*
- *direct training = practising with past exam papers*
- *specific training = studying a topic in depth*
- *TDTE cycle = test → diagnose the gap → train → evolve*
- *A2X Med = name of the mentorship programme*
- *medical school admission = the ultimate goal*

Writing style — critical for natural audio delivery:

1. *Use punctuation as rhythm and intention:*
    - *"..." → short pause / reflection*
    - *"—" → thought break, more natural than a comma*
    - *"," → micro pause*
    - *"." → full stop, new idea*
    - *"!" → genuine energy*
    - *"?" → genuine interrogative tone*
2. *Short sentences. Avoid long constructions — overly polished text sounds robotic in audio.*

3. *Use natural spoken contractions and rhythm: "you've", "that's", "let's", "what's", "didn't", "won't". Write as it sounds, not as formal prose.*

Rules:

- *Maximum 120 words*
- *Opens with an upbeat greeting: "Hey, [name]!" or "What's up, [name]!"*
- *Tone: informal, direct, authentic, energetic*
- *No false positivity — celebrate when earned, hold accountable when needed*
- *Write only the script, with no headings, bullet points, or formatting*

2) System prompt for active feedback, content clarification

Original version (Portuguese)

Você é um tutor acadêmico para estudantes que se preparam para o ENEM e vestibulares com foco em medicina. Seu papel é orientar o raciocínio do aluno sem fornecer respostas diretas — use o método socrático.

Quando o aluno apresentar uma dúvida de conteúdo (biologia, química, física, matemática, português, história, geografia ou qualquer outra disciplina do vestibular):

1. Pergunte primeiro o que ele já sabe sobre o tema: "O que você já sabe sobre isso?"
2. Faça perguntas que levem o aluno a descobrir a resposta por conta própria
3. Construa o raciocínio passo a passo, com andaimes ("scaffolding")
4. Nunca entregue a resposta diretamente — guie o aluno até ela
5. Seja conciso, claro e encorajador

Foque em raciocínio, interpretação e compreensão profunda, não em memorização. Responda sempre em português brasileiro.

English Adaptation

You are an academic tutor for students preparing for high-stakes university entrance exams with a focus on medical school admissions. Your role is to guide the student's reasoning without providing direct answers — use the Socratic method.

When a student presents a content question (biology, chemistry, physics, mathematics, Portuguese, history, geography, or any other subject area):

First ask what they already know about the topic: "What do you already know about this?"

Ask questions that lead the student to discover the answer on their own

Build reasoning step by step, using scaffolding

Never deliver the answer directly — guide the student toward it

Be concise, clear, and encouraging

Focus on reasoning, interpretation, and deep understanding, not memorisation. Always respond in English.

3) System prompt for routine and emotional support

Original version (Portuguese)

Você é um coach de aprendizagem para estudantes que se preparam para o ENEM e vestibulares com foco em medicina. Seu papel é apoiar o aluno com desafios não relacionados a conteúdo: rotina de estudos, ansiedade pré-vestibular, gestão do tempo, motivação e burnout.

REGRA PRINCIPAL — PRIMEIRA RESPOSTA:

Na sua PRIMEIRA mensagem, seja breve. Não liste ações, não dê conselhos ainda. Demonstre que você ouviu o aluno e faça 1 ou 2 perguntas abertas para entender melhor a situação dele antes de qualquer coisa. Só nas respostas seguintes, conforme o aluno compartilhar mais, você pode aprofundar e oferecer estratégias.

Diretrizes gerais:

1. Seja empático, acolhedor e não-clínico

2. Valide o estado emocional do aluno sem minimizar nem dramatizar

3. Ofereça estratégias baseadas em Autorregulação da Aprendizagem (SRL): planejamento, automonitoramento, estratégias de revisão — mas apenas quando já tiver contexto suficiente

4. Nunca diagnostique nem prescreva — você não é terapeuta nem psicólogo

5. Se o aluno descrever preocupações persistentes de saúde mental, sugira gentilmente conversar com um responsável ou profissional de saúde

Responda sempre em português brasileiro.

English Adaptation

You are a learning coach for students preparing for high-stakes university entrance exams with a focus on medical school admissions. Your role is to support students with non-content challenges: study routines, pre-exam anxiety, time management, motivation, and burnout.

MAIN RULE — FIRST RESPONSE:

In your FIRST message, be brief. Do not list actions or offer advice yet. Show that you have heard the student and ask 1 or 2 open questions to better understand their situation before anything else. Only in subsequent responses, as the student shares more, should you go deeper and offer strategies.

General guidelines:

Be empathetic, welcoming, and non-clinical

Validate the student's emotional state without minimising or dramatising

Offer strategies grounded in Self-Regulated Learning (SRL) — planning, self-monitoring, revision strategies — but only once you have sufficient context

Never diagnose or prescribe — you are not a therapist or psychologist

If the student describes persistent mental health concerns, gently suggest speaking with a trusted adult or mental health professional

Always respond in English.

B - CTAS-2 Survey

Original Version (Cassady & Johnson, 2014)

Cognitive Test Anxiety Scale – 2nd Edition

Please complete the following items using the four-point scale below.

1 = Not at all typical of me
2 = Somewhat typical of me
3 = Quite typical of me
4 = Very typical of me

1	I lose sleep over worrying about examinations.	1	2	3	4
2	I worry more about doing well on tests than I should.	1	2	3	4
3	I get distracted from studying for tests by thoughts of failing.	1	2	3	4
4	I have difficulty remembering what I studied for tests.	1	2	3	4
5	While preparing for a test, I often think that I am likely to fail.	1	2	3	4
6	I am not good at taking tests.	1	2	3	4
7	When I first get my copy of a test, it takes me a while to calm down to the point where I can begin to think straight.	1	2	3	4
8	At the beginning of a test, I am so nervous that I often can't think straight.	1	2	3	4
9	When I take a test that is difficult, I feel defeated before I even start.	1	2	3	4
10	While taking an important examination, I find myself wondering whether the other students are doing better than I am.	1	2	3	4
11	I tend to freeze up on things like intelligence tests and final exams.	1	2	3	4
12	During tests, I find myself thinking of the consequences of failing.	1	2	3	4
13	When I take a test, my nervousness causes me to make careless errors.	1	2	3	4
14	My mind goes blank when I am pressured for an answer on a test.	1	2	3	4
15	During tests, the thought frequently occurs to me that I may not be too bright.	1	2	3	4
16	During a course examination, I get so nervous that I forget facts I really know.	1	2	3	4
17	I do not perform well on tests.	1	2	3	4
18	During tests, I have the feeling that I am not doing well.	1	2	3	4
19	I am a poor test taker in the sense that my performance on a test does not show how much I really know about a topic.	1	2	3	4
20	After taking a test, I feel I should have done better than I actually did.	1	2	3	4
21	My test performances make me believe that I am not a good student.	1	2	3	4
22	I often realize mistakes I made right after turning in a test.	1	2	3	4
23	When I finish a hard test, I am afraid to see the score.	1	2	3	4
24	I don't seem to have much control over my test scores.	1	2	3	4

Portuguese Version (Silva, 2016)

Tabela 2

Médias, desvios-padrões e correlações item-total para cada item

N=279			
$CTAR_{25}$ itens	*M*	*DP*	Item-total R
1.Quando estou preocupado(a) com os testes/exames, perco o sono.	1,96	0,86	0,36
2.Preocupo-me com o meu desempenho nos testes, mais do que devia.	2,54	0,89	0,23
3.Quando estou a estudar para os testes sou distraído(a) por pensamentos acerca de poder falhar.	2,26	1,01	0,60
4.Tenho dificuldade em recordar aquilo que estudei para os testes.	2,00	0,88	0,55
5.Quando me estou a preparar ou a estudar para um teste, penso muitas vezes que provavelmente vou falhar.	2,16	1,01	0,72
6.Não sou bom(a) a fazer testes.	1,94	0,87	0,56
7.Quando recebo o enunciado do teste, demoro algum tempo a acalmar-me até conseguir começar a pensar com clareza.	1,81	0,91	0,46
8.No início de um teste, estou tão nervoso(a) que muitas vezes não consigo pensar com clareza.	1,99	0,94	0,65
9.Quando tenho um teste que é difícil, sinto-me derrotado(a) ainda antes de começar.	2,04	1,054	0,68
10.Durante um exame importante, dou por mim a questionar-me se os outros colegas estão a fazer melhor do que eu.	1,88	0,98	0,45
11.Tenho tendência a ficar bloqueado(a) em coisas como testes de inteligência, exames finais ou provas de avaliação.	2,14	0,94	0,64
12.Durante os testes, dou por mim a pensar quais serão as consequências do fracasso ou mau desempenho.	2,19	1,02	0,60
13.Quando estou a fazer um teste, o meu nervosismo leva-me a cometer enganos e pequenos erros.	2,59	0,96	0,63
14.Quando sou pressionado(a) para dar uma resposta num teste, a minha mente fica com uma “branca”.	2,43	0,98	0,67
15.Durante os testes ocorre-me frequentemente o pensamento de que posso não ser muito inteligente.	1,85	1,01	0,70
16.Durante um exame fico tão nervoso(a) que me esqueço de matéria que realmente sei.	1,89	0,89	0,61
17.Não me saio bem nos testes.	1,96	0,83	0,58
18.Durante os testes tenho a sensação que não estou a fazer bem o teste.	2,29	0,92	0,64
19.Sou fraco(a) nos testes, no sentido de que o meu desempenho num teste não mostra o quanto eu realmente sei sobre a matéria.	2,37	1,06	0,59
20.Depois de fazer um teste sinto que deveria ter feito melhor do que realmente fiz.	2,87	0,95	0,56
21.As minhas notas e o meu desempenho nos testes fazem-me acreditar que não sou um(a) bom(a) aluno(a).	2,07	1,05	0,65
22.É frequente aperceber-me de erros que fiz assim que acabo um teste.	2,82	0,96	0,32
23.Depois de fazer um teste difícil tenho medo de saber qual é a nota.	2,91	1,06	0,54
24.Quando tenho uma boa nota num teste, normalmente é porque tenho sorte.	1,59	0,84	0,30
25.Acho que não tenho muito controlo sobre as minhas notas nos testes.	1,85	0,84	0,55
Total	54,37	14,211	

Brazilian Portuguese Adaptation (author)

Portuguese Google Forms applied with students: https://forms.gle/hA7gPhN4TeQsHRhK6

1. Perco o sono me preocupando com provas e exames.
2. Me preocupo com meu desempenho nas provas mais do que deveria.
3. Quando estou estudando para uma prova, me distraio com pensamentos sobre a possibilidade de ir mal.
4. Tenho dificuldade em lembrar o que estudei para as provas.
5. Enquanto me preparo para uma prova, frequentemente penso que provavelmente vou ir mal.
6. Não sou bom(a) em fazer provas.
7. Quando recebo o caderno da prova, demoro um tempo para me acalmar até conseguir começar a pensar com clareza.
8. No início de uma prova, fico tão nervoso(a) que muitas vezes não consigo pensar com clareza.
9. Quando tenho uma prova difícil, me sinto derrotado(a) antes mesmo de começar.
10. Durante uma prova importante, me pego pensando se os outros colegas estão indo melhor do que eu.
11. Tenho tendência a travar em situações como testes de inteligência, exames finais ou provas de avaliação.
12. Durante as provas, me pego pensando nas consequências de reprovar ou ir mal.
13. Quando estou fazendo uma prova, meu nervosismo me faz cometer erros por descuido.
14. Minha mente dá um branco quando sou pressionado(a) a responder algo numa prova.
15. Durante as provas, frequentemente me ocorre o pensamento de que talvez eu não seja muito inteligente.
16. Durante um exame, fico tão nervoso(a) que esqueço conteúdos que realmente sei.
17. Não me saio bem nas provas.
18. Durante as provas, tenho a sensação de que não estou me saindo bem.
19. Sou fraco(a) em provas, no sentido de que meu desempenho não reflete o quanto realmente sei sobre o conteúdo.
20. Depois de fazer uma prova, sinto que deveria ter me saído melhor do que realmente fui.
21. Minhas notas e meu desempenho nas provas me fazem acreditar que não sou um(a) bom(a) aluno(a).
22. Com frequência percebo erros que cometi logo depois de entregar uma prova.
23. Quando termino uma prova difícil, tenho medo de ver a nota.
24. Não parece que tenho muito controle sobre minhas notas nas provas.
25. Com frequência percebo erros que cometi logo depois de entregar uma prova.

C - Interview Guide, Coding and Themes

Section 1 - Overall anxiety over study period

1. How did you perceive your anxiety over the last three weeks?
2. Was there something on the digital platform that helped you manage your anxiety? What? How?

Section 2 — Experience with the Feedback

1. Can you describe your overall experience using the feedback during your exam preparation?
2. How did you perceive the feedback you received (in terms of usefulness, clarity, and relevance)? Was there anything confusing or particularly helpful?

Section 3 — Emotional Impact (Cognitive Test Anxiety)

1. How did receiving the feedback affect how you felt about the upcoming exam? Did it increase or reduce your anxiety? How? Can you recall a specific moment?
2. Were there aspects of the feedback that made you feel more or less stressed or pressured? What exactly triggered that feeling? How did you respond to it?
3. Did the feedback affect your motivation or confidence while studying? In what way?

Section 4 — AI vs. Human Passive Feedback Perception

1. How did you experience receiving the audio feedback from [AI / a human]?
    a. Did you realize your feedback was AI generated? How?
2. Did it feel supportive or impersonal? Did you trust it? Why or why not?
3. Did the feedback made you study differently in the upcoming week?

Section 5 — AI vs. Human Active Feedback Perception

1. There was a “Help me!” button right on the platform dashboard, but you did (or did not) engage with it. Why? How did you feel about it?
2. How do you seek help during your studies? Does AI have a role in it? If so, how?

Section 6 — Barriers and Design Insights

1. What made it easier or harder for you to engage with the feedback during the preparation period? Were there moments you avoided it? Why?
2. What would you improve in the system?

Table

Coded Segments and Themes

Theme	Description	Coded Segments
AI as the Default Pragmatic Tool	Participants unanimously preferred AI for content-related doubts, citing speed, freedom to ask repeatedly, and absence of social judgment. AI was consistently identified as limited, however, when support required emotional attunement or relational depth.	AF – AI for content clarification (9) AF – AI advantages (6) AF – Teacher avoidance: content (6) AF – Teacher avoidance: emotional barriers (3) AF – Human is second option (2) PF – AI: automatic, impersonal (6) AF – AI for routine and organization (3) AF – AI for emotional regulation (2)
The Emotional Value of Human Presence	Human presence, whether in elaborate feedback, informal relationships, or simply as perceived availability, carried an irreducible emotional value that participants did not attribute to AI, alongside a paradoxical tendency to avoid actively seeking it.	PF – Real person looking after you (12) PF – Emotional regulation (3) PF – Perceives lower anxiety (2) AF – Help seeking avoidance in general (3) AF – AI avoidance: Fear of AI (2) AF – No help seeking for emotional regulation (1) Passive versus active feedback (1) AF – Peer feedback human (2) AF – Human connection (1) AF – Human: Possibility of help seeking, even if not used, is good (1) AF – Human: Ask questions during class (1)
Passive Feedback Enhances SRL and Motivation	The weekly passive feedback functioned as a catalyst for goal-directed behaviour, motivating study strategy adjustment and blurring the boundary between motivation and emotional regulation.	PF – Motivation (13) PF – Goal setting and SRL strategies (9) AF – AI + SRL Strategies (1)

Note. AF = Active Feedback codes; PF = Passive Feedback codes. Numbers in parentheses indicate coded segment frequency.

Non-exclusive licence to reproduce the thesis and make the thesis public

I , André von Eye Corleta,

1. grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis Supporting Cognitive Test Anxiety in Self-Regulated Learning: Comparing AI and Human Feedback in Digital Test Preparation, supervised by Prof. Natalia Edisherashvili;

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. am aware of the fact that the author retains the rights specified in points 1 and 2;

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

André von Eye Corleta

20/05/2026