ISRG Journal of Arts, Humanities and Social Sciences (ISRGJAHSS)

OPEN ACCESS

ISRG PUBLISHERS

Abbreviated Key Title: ISRG J Arts Humanit Soc Sci

ISSN: 2583-7672 (Online)

Journal homepage: https://isrgpublishers.com/isrgjahss

Volume – III Issue-I (January- February) 2025

Frequency: Bimonthly

Supervisory Feedback vs. AI: A Comparative Study on Postgraduate Student Satisfaction

Abdelouahd Bouzar[1*], Khaoula El Idrissi[2], Tayeb Ghourdou[3], Nirhech Ali[4]

[1,2,3] Department of English Studies, Sidi Mohamed Ben Abdellah University, Fes, Morocco

[4] Department of Linguistics, Massachusetts Amherst University, Massachusetts, United States

| **Received:** 07.02.2025 | **Accepted:** 11.02.2025 | **Published:** 13.02.2025

***Corresponding author:** Abdelouahd Bouzar

Department of English Studies, Sidi Mohamed Ben Abdellah University, Fes, Morocco

Abstract

Feedback plays a crucial role in postgraduate research supervision, influencing students' academic progress and satisfaction. Traditional supervisory feedback is valued for its engagement and contextual relevance, while artificial intelligence (AI)-generated feedback, particularly from models like ChatGPT, is gaining attention for its clarity and accessibility. However, limited quantitative research has explored students' comparative perceptions of AI versus human feedback. This study examines how postgraduate students perceive ChatGPT-generated feedback compared to traditional supervisory feedback. Specifically, it evaluates feedback clarity, relevance, accuracy, consistency, and comprehensiveness. Additionally, it investigates how different versions of ChatGPT (3.5 vs. 4) influence students' satisfaction with both AI-generated and supervisor-provided feedback. A cross-sectional study was conducted with 169 postgraduate students (M.A. and Ph.D.), who completed a structured questionnaire assessing their experiences with both feedback sources. Data were analyzed using independent t-tests to compare feedback perception across clarity, relevance and accuracy, and consistency and comprehensiveness. Supervisor feedback was also evaluated on engagement. Normality tests were performed before statistical analysis to ensure validity. Findings indicate that ChatGPT's feedback was rated higher in clarity than supervisor feedback, but lower in relevance, accuracy, and consistency. Supervisor feedback was perceived as more engaging and contextually appropriate. Users of ChatGPT 4 reported higher satisfaction with AI-generated feedback compared to ChatGPT 3.5 users, while supervisor feedback satisfaction remained consistent across both groups. Notably, reliance on AI for feedback increased as AI performance improved. AI-generated feedback offers advantages in accessibility and clarity but lacks the engagement and contextual awareness of human supervision. While AI tools can supplement academic guidance, they should not replace the critical thinking, personalized mentorship, and nuanced evaluation provided by human supervisors. The findings highlight the need for a balanced approach that integrates AI-driven feedback with traditional supervisory engagement to optimize postgraduate learning experiences.

Keywords: *Academic supervision, artificial intelligence, ChatGPT, feedback quality, postgraduate research, student perceptions.*

DOI: 10.5281/zenodo.14862530

Feedback in Academic Supervision

Definition and importance of feedback in academic research supervision

Feedback, a multifaceted concept with varying definitions, is pivotal across numerous professional fields, not as a monolith but as a nuanced, context-dependent mechanism. The essence of feedback lies in its communicative nature, as elucidated by Merriam-Webster, defining it as "*the transmission of evaluative or corrective information to its origin*" (Merriam-Webster, n.d.). "*It's the cyclical process where a system's output is harnessed to refine its input, enhancing performance in the technological realm, or steering self-corrective actions in automated control systems*" (Merriam-Webster, n.d.). In educational settings, feedback transcends the mere indication of correctness; it encompasses a spectrum of elements including precision, timeliness, guidance, and motivation, extending to advice on lesson sequencing (Mory, 2004, p. 745).

Hattie and Timperley (2007) argue that effective feedback addresses three cardinal questions: 'Where am I going?', 'How am I going?', and 'Where to next?', aligning respectively with feed up, feed back, and feed forward, thus guiding learners through a continuum of reflection and foresight (p. 88). Feedback's role as a pedagogical tool is further complexified when considering its types and the dynamism of its reception. Studies, such as those by Wentling (1973) and Hanna (1976), have demonstrated that the effectiveness of feedback is contingent on its form—partial or total—and the recipient's ability, with nuanced applications yielding different impacts on achievement and learning (as cited in Mory, 2004, p. 753).

The so-called "*feedback triad*" by Kulhavy and Wager (1993), referenced by Pereira et al. (2016b), encapsulates feedback's tripartite function: to motivate through response reinforcement, to inform by providing corrective avenues, and to reinforce by associating correct responses with prior stimuli (p. 746). This triad remains pertinent, emphasizing feedback's enduring importance. Phelps (2013) further categorizes feedback along several axes—direct or indirect, linear or interactional, formative or summative, and positive or corrective—each axis framing feedback in a light that suits the developmental trajectory of the recipient, whether it be a student or a professional in a supervisory relationship (p. 5).

Feedback in the context of teaching and learning

Feedback is a cornerstone of effective teaching and a catalyst for student learning enhancement. Pereira et al. (2016b) assert that it clarifies learning objectives and improvement strategies for students (p. 8). This is echoed by Mory (2004), who notes that feedback serves to affirm or modify student knowledge following practice or assessment activities (pp. 745–746). Historically, Sidney Pressey (as cited in Mory, 2004) underscored feedback's role in error correction as early as 1926, though subsequent research has debated its function, with some studies pointing to its punitive aspects regarding errors, resulting in inconsistent implications for learning (p. 746).

Further underlining its significance, feedback is integral to the assessment process in higher education, not merely as a corrective tool but as a driver of qualitative learning advancements and classroom dynamics (Mory, 2004, p. 7). In environments where formative assessment prevails, feedback is instrumental, helping students navigate and regulate their learning processes, consequently boosting academic achievement (Pereira et al., 2016b, p. 8). The value of feedback is also magnified when it is affirmative and learner-centered, fostering a more effective and relevant educational experience (Pereira et al., 2016b, p. 12). Reflecting on the early 20th-century research by Thorndike, Mory (2004) discusses that feedback, followed by a gratifying outcome, tends to reinforce behavior and promote learning (pp. 745–746).

Role of feedback in postgraduate research supervision

Al Bashir (2016) reiterates the latter sentiment, acknowledging feedback's critical role in enhancing higher education learning in general, and students' satisfaction in particular (p. 38). In postgraduate research supervision, student satisfaction emerges as a focal outcome, intricately tied to the quality of feedback provided. Ishak et al. (2021) define satisfaction as the favorable result of a task that bolsters an individual's self-esteem (p. 112), suggesting that the supervisory process significantly contributes to the student's personal and academic self-concept. This sentiment is echoed in their assertion that student satisfaction is fundamental for both development and motivation within educational settings (p. 112). Feedback, therefore, is not just a pedagogical tool but a determinant of satisfaction, with students yearning for constructive, timely discussions about their academic journey (Kushwah & Navrouzoglou, 2022, p. 55).

The supervisory practices and the ensuing feedback are pivotal in shaping this satisfaction. As described by Ishak et al. (2021), feedback within the supervisor-student dynamic should be an engagement rooted in professionalism and mutual respect (p. 109). The types of supervisory practices chosen can significantly influence not only the learning experience but also the satisfaction levels regarding the supervisory process (p. 111).

However, the provision and reception of feedback are laden with challenges that can impact satisfaction. Misalignments in expectations, cultural and linguistic barriers, and the inherently personal nature of research can complicate the feedback loop, potentially leading to dissatisfaction (East et al., 2012, p. 5). The ultimate aim of feedback in postgraduate supervision is to not only facilitate learning and project completion but also to engender a sense of satisfaction in students. This is achieved through feedback that is not only informative but also affirming, enhancing their self-esteem and motivation (Athiyaman, 1997, as cited in Ishak et al., 2021, p. 112). A harmonious supervisor-student relationship, where feedback is clear, constructive, and culturally sensitive, is essential for nurturing student satisfaction, which in turn is a key indicator of successful postgraduate research supervision.

ChatGpt's Vs Supervisor's Feedback

The review of literature, concerning feedback from supervisors, highlights key areas critical in influencing students' progression in their research endeavours. These areas include the clarity and adaptability of feedback, its supportive and developmental nature, and the level of interaction and engagement it fosters. In examining ChatGPT's feedback, a different perspective is adopted. The interactivity, engagement, and adaptability of ChatGPT's responses largely hinge on the user's prompts, presenting a contrast to supervisor feedback, which is more influenced by the personal attributes of the supervisor and less by the supervisee's input. However, the clarity, relevance, accuracy, consistency, and comprehensiveness of ChatGPT's feedback emerge as contentious points in the literature, raising questions that are less dependent on user prompts and more on the inherent capabilities of the AI system.

DOI: 10.5281/zenodo.14862530

Supervisors's Feedback

Clear and Adaptable

In the context of postgraduate supervision, the efficacy of feedback hinges significantly on its clarity and adaptability. These two criteria form the cornerstone of effective communication between supervisors and students, enhancing the overall learning experience. Clarity in feedback ensures that the guidance provided is easily understandable and unambiguous, directly addressing the student's work. Adaptability, on the other hand, refers to tailoring feedback to suit the individual needs and contexts of each student, thereby making it more relevant and impactful.

The concept of clarity in feedback is multifaceted. It involves the precision of language, the relevance of the content, and the specificity of the guidance provided. Clear feedback directly addresses the key aspects of a student's work, making it easy for them to understand what is expected of them and how they can improve. Duncanson et al. (2020) emphasize the importance of regular and clear communication in this regard, where feedback is instructive and tailored to the specific task, thereby ensuring its relevance and applicability (p. 17).

Adaptability in feedback is equally crucial. It involves understanding the unique challenges and strengths of each student and adjusting the feedback accordingly. This personalized approach, as highlighted by East et al. (2012, p. 15), fosters a more engaging and collaborative partnership in the supervisory relationship. By adapting feedback to the specific needs and contexts of students, supervisors can ensure that the guidance provided is not only relevant but also resonates with the individual learning trajectories of the students.

Supportive

The role of a supportive supervisory relationship in postgraduate education is integral to enhancing student engagement and performance. Supervisors who consistently offer support and reassurance, as noted by Ishak et al. (2021, p. 111), play a significant role in boosting student motivation, thus facilitating progress in research endeavors. The support provided in these relationships extends beyond basic encouragement. It involves a deeper appreciation of students' ideas, bolstering their confidence, and promoting active engagement in the learning process. This approach aligns well with specific educational objectives, as detailed by Khuram et al. (2023, p. 3), and is crucial in nurturing academic skills. Furthermore, it fosters creativity and a drive for knowledge exploration, which are essential for enhanced research performance and productivity.

The interactive nature of feedback within these relationships is also pivotal in fostering an environment conducive to reflection and growth. Duncanson et al. (2020, p. 14) emphasize the importance of engaging in a dialogue that includes justifications for comments and suggestions. Such a dialogue helps in building a constructive environment where progress is grounded in understanding and reflection.

In this context, the role of the supervisor in providing constructive feedback becomes essential. This feedback should be aimed at encouraging students to delve deeply into their work, fostering a sense of accomplishment and belonging within their academic community. Such a sense of belonging contributes significantly to student satisfaction, as observed by Bastola & Hu (2021, p. 1) and East et al. (2012, p. 3).

Moreover, the nature of the advising relationship has broader implications in the academic sphere. Windon (2020, p. 264) highlights that a positive supervisory relationship contributes to a supportive departmental environment, successful integration into the academic community, and timely degree completion. In contrast, a poor relationship can lead to adverse outcomes, including the possibility of students abandoning their doctoral studies.

The manner in which feedback is delivered in these relationships is critical. Bastola & Hu (2021, p. 9) note the importance of providing critiques in a balanced and positive manner within a supportive framework, which can significantly enhance the educational experience. They also emphasize the necessity of tailoring feedback to the specific needs of students, using evaluative language with care, and providing a balanced perspective that acknowledges strengths and areas for improvement. Such a balanced approach ensures that feedback is both relevant and constructive, catering to the unique requirements of each student. This approach underscores the importance of constructive feedback in fostering an environment conducive to academic and personal development.

Engaging and Interactive

In postgraduate supervision, the interaction and engagement of feedback are crucial elements. The process of regular supervisor-student meetings and interactive dialogues, as described by East et al. (2012, p. 15), transcends the limitations of one-way feedback. This method fosters an iterative cycle of shared interpretations and clarified expectations, enhancing the educational process. The engagement of both parties in this process is vital for effective communication and understanding.

The manner in which feedback is delivered is also a key factor in creating an engaging learning environment. According to Khuram et al. (2023, p. 3), feedback should be delivered in a way that not only engages students but also encourages their active participation in the learning process. This approach involves an exchange of ideas and receipt of constructive feedback, which is instrumental in promoting students' academic and professional development.

ChatGpt's Feedback

Clarity

The evolving capabilities of ChatGPT in providing personalized and effective feedback in educational contexts have garnered significant attention. AlAfnan et al. (2023) critically assessed ChatGPT, highlighting that while responses were generally accurate, they often lacked structured clarity and sometimes replicated content without appropriate referencing (p. 63). Borji (2023) extended this critique, emphasizing ChatGPT's challenges, including the generation of imprecise information, potential biases, and inconsistencies in clarity and reliability.

In contrast, Javaid et al. (2023) and Baskara (2023) recognized ChatGPT's capacity for enhancing feedback through personalization. They noted that ChatGPT could adapt its feedback to individual students' writing styles and needs, making the guidance clear, instructive, and tailored (Javaid et al., 2023, p. 2; Baskara, 2023, p. 98). This personalization ensures that the feedback is not only understandable but also directly applicable to the student's specific context.

Supporting this perspective, Jacobsen & Weber (2023) found that ChatGPT's feedback quality often parallels expert feedback and exceeds that of novices, showcasing its precision and relevance (p.

DOI: 10.5281/zenodo.14862530

24). Furthermore, Escalante et al. (2023) observed that while there is variability in agreement between AI and human feedback, ChatGPT typically offers more detailed and readable feedback, further emphasizing its clarity (p. 4). These insights collectively highlight the potential of ChatGPT as a tool for enhancing the quality and specificity of feedback in educational settings.

Relevance & Accuracy/correctness

ChatGPT's performance in providing responses to life support exams at a university was found to be not only relevant and accurate but also more congruent with resuscitation guidelines compared to previous AI tools, posing a significant challenge to academic integrity in higher education (Cotton et al., 2023; Fijačko et al., 2023). Its proficiency in generating contextually relevant replies stems from its language modeling capabilities, where it predicts the likelihood of each word based on preceding ones, resulting in syntactically and grammatically coherent responses (Shihab et al., 2023).

Furthermore, with access to appropriate diagnostic data, ChatGPT has shown an accuracy of 76.9% in making correct diagnoses, underscoring its effectiveness in clinical settings (Rao et al., 2023). Its ability to assimilate and learn from vast data sets enables it to deliver highly relevant and accurate answers to diverse queries (Shihab et al., 2023). Developed by OpenAI, ChatGPT's architecture, based on the Generative Pre-trained Transformer (GPT), facilitates this learning process and the generation of contextually appropriate responses (Shihab et al., 2023). These capabilities are continually evolving, making ChatGPT models reliable sources for a wide array of clinical inquiries (King et al., 2023).

The capabilities of Large Language Models (LLMs) like ChatGPT in processing and responding to complex queries showcase significant advantages. LLMs have demonstrated remarkable proficiency in mimicking human language, enabling them to understand complex inquiries and provide relevant, real-time answers (Deng & Lin, 2023; Biswas, 2023). Their ability to process vast amounts of medical literature and deliver contextually relevant information makes them valuable for educators and students (Dhanvijay et al., 2023). In certain studies, ChatGPT has successfully generated factually accurate and contextually appropriate responses to intricate clinical questions (Li et al., 2023). This capability to rapidly produce sophisticated and accurate responses, aligned with human needs and desires, has elicited substantial interest and excitement in various fields (Li et al., 2023; Shihab et al., 2023).

Conversely, the limitations of LLMs pose significant challenges. These models, including ChatGPT, can generate biased, offensive, or incorrect responses, especially to open-ended questions (Hsu et al., 2023). They may struggle with the nuanced uses of human language such as sarcasm, irony, and humor, leading to inappropriate or irrelevant reactions (Shihab et al., 2023). There's a notable risk of "hallucinations," where the model generates irrelevant or nonsensical responses (Hsu et al., 2023). In education and medical contexts, these limitations can be particularly problematic. For example, inaccuracies in complex patient scenarios or drug-related questions can lead to misinformation or treatment risks (Al-Dujaili et al., 2023; Hsu et al., 2023). Additionally, GPT-3.5 has shown to produce more verbose and inaccurate responses compared to its successor, GPT-4, highlighting the ongoing need for improvement in accuracy and reliability (King et al., 2023 ; Chen et al., 2023).

Consistency and comprehensiveness

ChatGPT, particularly in its GPT-4 iteration, offers significant workload reduction and faster response times, efficiently addressing complex queries with prompt and comprehensive answers (Koubaa et al., 2023). Compared to its predecessor GPT-3.5, GPT-4 exhibits a notable increase in the frequency of providing detailed and coherent responses (King et al., 2023; Al-Dujaili et al., 2023; AlAfnan et al., 2023). Moreover, this version shows enhanced language comprehension skills, especially in interpreting negative expressions and antonyms, marking a clear improvement over earlier versions (Abujaber et al., 2023).

Despite these advancements, several challenges persist. ChatGPT is prone to inconsistencies and contradictions, particularly in logical predictions and paraphrased content, often leading to contradictory outputs (Jang & Lukasiewicz, 2023; Abujaber et al., 2023). It also struggles with tasks requiring original thinking or responses to open-ended questions (Shihab et al., 2023). Additionally, its ability to generalize to new, unseen data remains limited, affecting its applicability in diverse real-world contexts (Koubaa et al., 2023).

Another limitation is its surface-level understanding, as ChatGPT primarily processes information based on pattern recognition, sometimes resulting in off-topic or superficial responses (Farrokhnia et al., 2023). The model also shows less competence in higher-order thinking skills, such as critical and analytical thinking (Farrokhnia et al., 2023). Lastly, concerns regarding the consistency and unity of generated content have been raised, with the model sometimes altering its decisions based on paraphrased inputs and producing incoherent paragraphs (Abujaber et al., 2023; AlAfnan et al., 2023).

Negative supervisory experiences and Reliance on ChatGpt

While the interactive approach to feedback in postgraduate supervision, particularly with the aid of tools like ChatGPT, has its benefits in enhancing the learning experience, it is crucial to consider the other side of this equation. Negative supervisory experiences in postgraduate research can have significant and far-reaching consequences for students, impacting not only their satisfaction levels but also their confidence and overall academic progress. Ishak et al. (2021) point out that issues like communication problems and conflicts during supervision can lead to reduced satisfaction with the supervisor and result in less confident, open, and experimental supervisees, as further discussed by Kushwah & Navrouzoglou (2022) and Bastola & Hu (2021).

This dissatisfaction often extends to core research areas. Students may struggle with various aspects of their research, feeling unsupported in improving their academic writing or dealing with challenges in selecting research topics, reviewing literature, developing research designs, analyzing data, and interpreting results. The quantity and quality of feedback received in these areas are often sources of dissatisfaction, as noted by Bastola & Hu (2021).

Moreover, a perceived lack of supervisory support can lead to feelings of inadequacy, low self-esteem, and reduced agency among students. The absence of timely and comprehensive feedback from supervisors exacerbates these issues, leaving students feeling neglected and discouraged. Pereira et al. (2016b) echo these sentiments, noting challenges in higher education feedback, including students' dissatisfaction with comments on

DOI: 10.5281/zenodo.14862530

their work and the timeliness of feedback. Even teaching staff find providing feedback burdensome and question its effectiveness.

These findings underscore the critical role of effective, supportive supervision in postgraduate research and the detrimental effects that negative supervisory experiences can have on student satisfaction, confidence, and academic achievement. In this context, the reliance on ChatGPT for feedback, despite its advantages, poses several risks. Students may become overly dependent on AI, as cautioned by Qadir (2022), potentially hindering their ability to take ownership of their learning. Additionally, issues such as data security, algorithmic bias, and negative effects on socialization and collaboration among students, as highlighted by Baskara (2023), present significant challenges. Furthermore, while ChatGPT excels in generating text, it lacks the capability to address the emotional and psychological needs of students, an area where human teachers play a crucial role in providing support and fostering a positive, inclusive classroom environment, as noted by Atlas (2023).

Overall, ChatGPT offers a fast, detailed, and personalized feedback mechanism that can supplement traditional supervisory feedback. However, its limitations, particularly in addressing the emotional and psychological aspects of learning and the risk of overreliance, highlight the importance of maintaining a balanced approach that leverages the strengths of both AI and human interaction in educational settings.

Purpose of the Study and Research Questions

In the comparative analysis of ChatGPT and traditional supervisory feedback, an important gap in the literature is the lack of quantitative studies and a focus primarily on supervisors' perspectives. Most existing research, being qualitative, provides insights but lacks definitive, numerically backed findings. Additionally, there's a notable shortage of studies centered on students' perceptions of feedback, crucial for student-centered learning.

Understanding how students perceive and are affected by feedback, whether from AI tools like ChatGPT or human supervisors, is essential. This gap in the literature suggests a need for more balanced research that includes both quantitative analysis and a focus on students' experiences and needs, particularly in the context of thesis writing and academic development. Addressing this gap could lead to more effective feedback practices in educational settings, combining the rapid, detailed responses of AI with the nuanced, empathetic approach of human supervision. For this purpose, this research to answer the following questions:

(1) How do students perceive ChatGPT's feedback compared to a supervisor's in terms of clarity, consistency, and relevance and accuracy?

(2) How does the version of ChatGPT (3.5 vs. 4) influence the perceived accuracy, clarity, and consistency and relevance of supervisory and ChatGPT's feedback?

(3) To what extent does the version of ChatGPT (3.5 vs. V4) influence users' overall satisfaction with supervisor-provided feedback and ChatGPT-generated feedback?

Methodology

Research Design

A cross-sectional methodological framework was utilized to examine students' perceptions of supervisor and ChatGPT feedback, focusing mainly on M.A and Ph.D. students. An independent t-test was conducted as the primary statistical analysis to address the main research questions. This methodology was chosen to identify significant differences in feedback perception across various dimensions, such as clarity, consistency, and relevance and accuracy.

Participants

The study involved 169 participants, all whom are graduate students (table 1). MA students comprised the larger proportion (n = 113), compared to PhD students (n = 56). The distribution across age groups indicated that the majority fell within the 26–30-year range (n = 71), followed by participants aged 36+ (n = 42), while those aged 18–25 and 31–35 each constituted 16.6% (n = 28). Gender distribution skewed predominantly toward females (n = 126), with males representing 25.4% (n = 43) of the sample. Regarding tool usage, ChatGPT 3.5 was employed by most participants (n = 140), while ChatGPT 4 was used by a smaller subset (n = 29, 17.2%). All variables demonstrated complete data (Valid N = 169), with no missing responses.

Table 1: Participant Demographics

Variable	Category	Frequency	Valid N
Age	18-25	28	169
	26-30	71	
	31-35	28	
	36+	42	
Gender	Males	43	
	Females	126	
Academic Level	PhD students	56	
	MA students	113	
ChatGPT version	ChatGPT 3.5	140	
	ChatGPT 4	29	

Data collection

Data were collected through a structured questionnaire, which yielded a total of 169 responses. The questionnaire was distributed during formal M.A. classes and shared via social media platforms such as WhatsApp and Facebook groups specifically designed for Ph.D. and M.A. students. Additionally, professors, Ph.D. candidates, and M.A. students assisted in disseminating and supervising the questionnaire to ensure a diverse pool of respondents. To further enhance participant diversity, 70 responses were obtained through Prolific, an online platform that facilitates research participation across varied demographic backgrounds.

The questionnaire comprised four sections: (1) Demographics, (2) Experience with Supervisory Feedback, (3) Satisfaction with Supervisor's Feedback, and (4) Satisfaction with ChatGPT's Feedback. The first section gathered basic demographic information, including age, gender, and academic level. The second section functioned as a branching filter, ensuring that only participants who had experience with both ChatGPT and supervisor feedback proceeded to the subsequent sections; those who had never used ChatGPT or had not received supervisor feedback were excluded from further participation.

DOI: 10.5281/zenodo.14862530

The third and fourth sections measured participants' satisfaction with feedback received from their supervisor and ChatGPT, respectively. Respondents rated clarity, relevance and accuracy, consistency and comprehensiveness for both sources of feedback. Given the human nature of supervisory feedback, the supervisor evaluation section included one additional dimension: engagement. Satisfaction was assessed using statements such as *"I am satisfied with the supervisor's/ChatGPT's ability to provide efficient feedback in one session without the need for follow-ups," "I am satisfied with the supervisor's/ChatGPT's ability to follow my research directions," "I am satisfied with the coherence of the supervisor's/ChatGPT's feedback," "I am satisfied with the supervisor's adequacy in addressing grammar and style,"* and *"I am satisfied with the supervisor's ease of interaction and quick response."*

These statements corresponded to the main key dimensions of feedback evaluation: clarity, relevance and accuracy, consistency and comprehensiveness. The structured nature of the questionnaire ensured a systematic and comprehensive assessment of participants' experiences with both sources of feedback.

Data analysis

Data were analyzed using SPSS (Statistical Package for the Social Sciences) as the primary statistical software. To address the research questions, an independent samples t-test was conducted to compare participants' perceptions of feedback across different conditions.

Scales

The reliability analysis (table 2) reveals that the scales used in this study demonstrate acceptable to excellent internal consistency. Supervisor's Feedback scales show Cronbach's Alpha values ranging from 0.710 for Consistency and Comprehensiveness (2 items) to 0.961 for Engagement (4 items), indicating that the items within each scale consistently measure the intended construct despite varying item counts. Similarly, ChatGPT's Feedback scales yield Alpha coefficients of 0.802 for Clarity, 0.845 for Relevance and Accuracy, and 0.827 for Consistency and Comprehensiveness, suggesting reliable performance across the dimensions. Although the slightly lower Alpha for Supervisor's Feedback Consistency and Comprehensiveness may be influenced by its fewer items, it remains within acceptable reliability standards. Overall, these findings support the robustness of the measurement instruments and lend confidence to subsequent analyses based on these scales.

Table 2: Reliability Analysis of Feedback Scales

Scale	Cronbach's Alpha	N of Items
Supervisor's Feedback Clarity	0.844	3
Supervisor's Feedback Relevance and Accuracy	0.839	3
Supervisor's Feedback Consistency and Comprehensiveness	0.71	2
Supervisor's Feedback Engagement	0.961	4
ChatGPT's Feedback Clarity	0.802	3
ChatGPT's Feedback Relevance and Accuracy	0.845	3
ChatGPT's Feedback Consistency and Comprehensiveness	0.827	2

Normality Test

Before conducting the independent samples t-test, a normality test was performed to determine whether the data followed a normal distribution, ensuring the appropriateness of parametric statistical tests. Both the Kolmogorov-Smirnov (K-S) test and the Shapiro-Wilk (S-W) test (table 3) were applied to assess the distribution of the variables. The results indicated that all tested variables—Supervisor Clarity, Supervisor Relevance and Accuracy, Supervisor Engagement, ChatGPT Relevance and Accuracy, ChatGPT Consistency and Comprehensiveness, and ChatGPT Clarity—yielded significant p-values ($p < .001$) in both tests, suggesting a significant deviation from normality.

The Shapiro-Wilk test, which is particularly appropriate for small sample sizes, also produced highly significant values ($p < .001$) for all variables, confirming the non-normal distribution of the data. Given these findings, normality violations were considered before proceeding with the independent t-test, and the robustness of the t-test to deviations from normality was taken into account when interpreting the results.

Table 3: Normality Test Results for Feedback Variables

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Supervisor's Feedback Clarity	.337	28	.000	.639	28	.000
Supervisor's Feedback Relevance and Accuracy	.337	28	.000	.639	28	.000
Supervisor's Feedback Consistency and Comprehensiveness	.	28	.	.	28	.
Supervisor's Feedback Engagement	.337	28	.000	.639	28	.000
ChatGPT's Feedback Clarity	.337	28	.000	.639	28	.000
ChatGPT's Feedback Relevance and Accuracy	.337	28	.000	.639	28	.000
ChatGPT's Feedback Consistency and Comprehensiveness	.337	28	.000	.639	28	.000

DOI: 10.5281/zenodo.14862530

Results

(1) How do students perceive ChatGPT's feedback compared to a supervisor's in terms of clarity, consistency, and relevance and accuracy?

Supervisor's and ChatGPT's feedback were evaluated on comparable dimensions of clarity, relevance and accuracy, and consistency and comprehensiveness, with Supervisor's feedback also assessed on engagement. Regarding clarity, ChatGPT's feedback yielded a slightly higher mean difference (5.16174) than that of Supervisor's (5.04931) (table 4). In contrast, Supervisor's feedback demonstrated higher mean differences for both relevance and accuracy (5.44444 versus 4.82840) and for consistency and comprehensiveness (5.37500 versus 4.74260) compared to ChatGPT's. Additionally, Supervisor's feedback on engagement recorded a mean difference of 4.50442. All differences were statistically significant (p = 0.000).

Table 4: Comparative Mean Differences and Significance of Feedback Variables

	Mean Difference	Sig. (2-tailed)
Supervisor's Feedback Clarity	5.04931	0.000*
ChatGPT's Feedback Clarity	5.16174	0.000*
Supervisor's Feedback Relevance and Accuracy	5.44444	0.000*
ChatGPT's Feedback Relevance and Accuracy	4.82840	0.000*
Supervisor's Feedback Consistency and Comprehensiveness	5.37500	0.000*
ChatGPT's Feedback Consistency and Comprehensiveness	4.74260	0.000*

(2) How does the version of ChatGPT (3.5 vs. 4) influence the perceived accuracy, clarity, and consistency and relevance of supervisory and ChatGPT's feedback?

Table 5: Impact of ChatGPT Versions on Feedback Perception

	ChatGPT Version		Sign. (2-tailed)
	Chatgpt 3.5	Chatgpt V4	
Supervisor's Feedback Clarity	5.1333	0.000*	0.000*
Supervisor's Feedback Relevance and Accuracy	5.5333	5.0000	0.016*
Supervisor's Feedback Relevance and Accuracy	5.5500	4.5000	0.000*
ChatGPT's Feedback Clarity	5.0333	5.7816	0.004*
Supervisor's Feedback Consistency and Comprehensiveness	4.7667	5.1264	0.016*
ChatGPT's Feedback Clarity	4.5500	5.6724	0.066

The data (table 5) indicate differences in perceptions of both supervisory and ChatGPT-generated feedback based on the ChatGPT version used by participants. Specifically, for supervisory feedback, users of ChatGPT 3.5 (n = 140) reported higher ratings in clarity (M = 5.1333, p = 0.000), relevance and accuracy (M = 5.5333, p = 0.016), and consistency and comprehensiveness (M = 5.5500, p = 0.000) compared to users of ChatGPT V4 (n = 29 for clarity; n = 28 for relevance and accuracy and consistency and comprehensiveness) who reported mean ratings of 4.6437, 5.0000, and 4.5000, respectively. In contrast, when evaluating ChatGPT's own feedback, users of ChatGPT V4 provided higher ratings for clarity (M = 5.7816, p = 0.004) and relevance and accuracy (M = 5.1264, p = 0.016) than those using ChatGPT 3.5 (n = 140; M = 5.0333 for clarity and M = 4.7667 for relevance and accuracy). Although the difference in ratings for the consistency and comprehensiveness of ChatGPT's feedback between ChatGPT 3.5 users (M = 4.5500) and ChatGPT V4 users (M = 5.6724) approached conventional significance (p = 0.066), it did not reach the standard threshold.

(1) To what extent does the version of ChatGPT (3.5 vs. V4) influence users' overall satisfaction with supervisor-provided feedback and ChatGPT-generated feedback?

For overall satisfaction with feedback (table 6), the data reveal differential outcomes based on the ChatGPT version used. Specifically, regarding supervisor's feedback, ChatGPT 3.5 users (n = 84) reported a mean of 4.8750, compared to a mean of 4.6667 among ChatGPT V4 users (n = 28), a difference that was not statistically significant (p = 0.292). In contrast, for ChatGPT's own feedback, a significant difference was observed; ChatGPT V4 users (n = 29) reported a higher mean satisfaction (5.5268) than ChatGPT 3.5 users (n = 140), with the difference reaching statistical significance (p = 0.003).

DOI: 10.5281/zenodo.14862530

Table 6: Comparison of Overall Satisfaction with Supervisor and ChatGPT Feedback Across Versions

	ChatGPT Version		Sign. (2-tailed)
	Chatgpt 3.5	Chatgpt V4	
Overall Satisfaction with Supervisor's Feedback	4.8750	4.6667	0.292
Overall Satisfaction with ChatGPT's Feedback	4.7833	5.5268	0.003*

Discussion

This study examined students' perceptions of ChatGPT-generated feedback in comparison to supervisor feedback across key dimensions, including clarity, relevance, accuracy, and consistency. Additionally, it investigated how the version of ChatGPT (3.5 vs. 4) influenced these perceptions and overall satisfaction with feedback. The findings contribute to the growing body of research on AI-generated feedback and its implications for academic supervision.

Perceptions of ChatGPT's Feedback Compared to Supervisor Feedback

The results indicate that students perceive ChatGPT's feedback as slightly clearer than supervisor feedback, with ChatGPT recording a marginally higher mean score compared to supervisors. This aligns with existing literature, which highlights AI's ability to provide structured and immediate responses that minimize ambiguity (Jacobsen & Weber, 2023). However, despite its clarity, ChatGPT's feedback was rated lower than supervisors' in terms of relevance and accuracy and consistency and comprehensiveness. These findings corroborate earlier studies that caution against AI-generated content's potential for factual inaccuracies, biases, and inconsistency in responses (Borji, 2023; AlAfnan et al., 2023).

Supervisor feedback, while slightly lower in clarity, was perceived as more reliable and consistent. This can be attributed to the expertise, contextual awareness, and critical judgment that human supervisors bring to feedback, which AI systems often struggle to replicate (Escalante et al., 2023). Moreover, the significantly higher rating for supervisor engagement suggests that students value human interaction and personalized guidance, an element largely absent in AI-generated feedback. This finding is consistent with prior research emphasizing the role of supervisor-student relationships in fostering academic development and motivation (Duncanson et al., 2020; Bastola & Hu, 2021).

Impact of ChatGPT Version on Perceptions of Feedback

The study further explored whether students' perceptions of feedback varied based on the ChatGPT version used. The results indicate that ChatGPT 3.5 users rated supervisor feedback higher in clarity, relevance and accuracy, and consistency and comprehensiveness compared to ChatGPT V4 users. This suggests that students using the older version of ChatGPT placed greater reliance on human supervision, potentially due to the recognized limitations of ChatGPT 3.5 in generating reliable and structured feedback (King et al., 2023).

Conversely, ChatGPT V4 users rated ChatGPT's feedback higher in clarity, relevance and accuracy, and consistency and comprehensiveness. These results are in line with existing research, which highlights the improved performance of GPT-4 in providing more coherent, contextually relevant, and accurate responses compared to GPT-3.5 (Al-Dujaili et al., 2023; Koubaa et al., 2023). While these findings suggest that advancements in AI technology enhance feedback quality, it is noteworthy that students using GPT-4 were less satisfied with supervisor feedback, potentially reflecting a shift in expectations as AI feedback quality improves.

Overall Satisfaction with Supervisor and ChatGPT Feedback

The results on overall satisfaction with feedback highlight a significant divergence between ChatGPT and human supervisors. While no significant difference was observed in students' satisfaction with supervisor feedback between ChatGPT 3.5 and ChatGPT V4 users, satisfaction with ChatGPT's feedback was significantly higher among ChatGPT V4 users. This aligns with the increasing trend of AI reliance in academic writing and research assistance, where improved AI models are perceived as more effective and informative (Javaid et al., 2023).

However, this increasing satisfaction with ChatGPT feedback raises concerns about the potential overreliance on AI-generated feedback at the expense of human interaction. Prior studies suggest that while AI feedback is useful, it lacks the contextual awareness, critical thinking, and personal engagement that human supervisors provide (Pereira et al., 2016b; Kushwah & Navrouzoglou, 2022). Overdependence on ChatGPT may also limit students' ability to engage in meaningful academic discourse and develop self-regulatory learning strategies, as warned by Qadir (2022).

Implications and Future Considerations

These findings have several implications for academic supervision and the integration of AI tools in higher education. First, while AI models such as ChatGPT can enhance feedback accessibility and clarity, they should be used as complementary tools rather than replacements for human supervision. The significantly higher ratings for supervisor feedback in relevance, accuracy, and consistency underscore the irreplaceable role of human expertise in guiding students.

Second, institutions should recognize the potential influence of AI model updates on students' expectations and satisfaction with feedback. As AI-generated responses become more structured and accurate, students may develop higher expectations for feedback quality, potentially leading to dissatisfaction with traditional supervisory approaches. Supervisors should adapt to these shifts by incorporating AI-generated feedback in their mentoring processes while ensuring that students remain engaged in human-led academic discussions.

Finally, future research should explore the long-term impact of AI feedback reliance on academic writing quality, student learning behaviors, and the supervisor-student dynamic. A more nuanced investigation into how AI tools can be optimally integrated into academic supervision without diminishing the value of human engagement is necessary.

Conclusion

This study provides valuable insights into how students perceive feedback from ChatGPT and human supervisors, highlighting the strengths and limitations of AI-generated feedback. While

DOI: 10.5281/zenodo.14862530

ChatGPT demonstrates strong performance in clarity and is improving in accuracy and consistency, supervisor feedback remains superior in providing relevant, accurate, and engaging guidance. The differences observed between ChatGPT 3.5 and V4 users suggest that as AI models improve, students' expectations for feedback evolve accordingly. These findings emphasize the need for a balanced approach that leverages AI's efficiency while preserving the depth, context, and engagement of human supervision.

References

1. Abujaber, A. A., Abd-alrazaq, A., Al-Qudimat, A. R., & Nashwan, A. J. (2023). A Strengths, Weaknesses, Opportunities, and Threats (SWOT) Analysis of ChatGPT Integration in Nursing Education: A Narrative Review. *Cureus*. https://doi.org/10.7759/cureus.48643
2. Al Bashir, M. (2016). The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education. *Journal of Education and Practice*, *7*(16), 38–41.
3. AlAfnan, M. A., Samira Dishari, Marina Jovic, & Koba Lomidze. (2023). ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses. *Journal of Artificial Intelligence and Technology*. https://doi.org/10.37965/jait.2023.0184
4. Al-Dujaili, Z., Omari, S., Pillai, J., & Al Faraj, A. (2023). Assessing the accuracy and consistency of ChatGPT in clinical pharmacy management: A preliminary analysis with clinical pharmacy experts worldwide. *Research in Social and Administrative Pharmacy*, *19*(12), 1590–1594. https://doi.org/10.1016/j.sapharm.2023.08.012
5. Atlas, S. (2023). *ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI*. The University of Rhode Island. https://digitalcommons.uri.edu/cba_facpubs/548?utm_source=digitalcommons.uri.edu%2Fcba_facpubs%2F548&utm_medium=PDF&utm_campaign=PDFCoverPages
6. Baskara, Fx. R. (2023). The Promises and Pitfalls of Using Chat GPT for Self-Determined Learning in Higher Education: An Argumentative Review. *Prosiding Seminar Nasional Fakultas Tarbiyah Dan Ilmu Keguruan IAIM Sinjai*, *2*, 95–101. https://doi.org/10.47435/sentikjar.v2i0.1825
7. Bastola, M. N., & Hu, G. (2021). "Chasing my supervisor all day long like a hungry child seeking her mother!": Students' perceptions of supervisory feedback. *Studies in Educational Evaluation*, *70*, 101055. https://doi.org/10.1016/j.stueduc.2021.101055
8. Biswas, S. (2023). ChatGPT and the Future of Medical Writing. *Radiology*, *307*(2), e223312. https://doi.org/10.1148/radiol.223312
9. Borji, A. (2023). *A Categorical Archive of ChatGPT Failures*. https://doi.org/10.48550/ARXIV.2302.03494
10. Chen, L., Zaharia, M., & Zou, J. (2023). *How is ChatGPT's behavior changing over time?* (arXiv:2307.09009). arXiv. http://arxiv.org/abs/2307.09009
11. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1–12. https://doi.org/10.1080/14703297.2023.2190148
12. Deng, J., & Lin, Y. (2023). The Benefits and Challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81–83. https://doi.org/10.54097/fcis.v2i2.4465
13. Dhanvijay, A. K. D., Pinjar, M. J., Dhokane, N., Sorte, S. R., Kumari, A., & Mondal, H. (2023). Performance of Large Language Models (ChatGPT, Bing Search, and Google Bard) in Solving Case Vignettes in Physiology. *Cureus*. https://doi.org/10.7759/cureus.42972
14. Duncanson, K., Schmidt, D., & Webster, E. (2020). Giving and receiving written feedback on research reports: A narrative review and guidance for supervisors and students. *Health Education in Practice: Journal of Research for Professional Learning*, *3*(2). https://doi.org/10.33966/hepj.3.2.14767
15. East, M., Bitchener, J., AUT University, Basturkmen, H., & The University of Auckland. (2012). What constitutes effective feedback to postgraduate research students? The students' perspective. *Journal of University Teaching and Learning Practice*, *9*(2), 103–119. https://doi.org/10.53761/1.9.2.7
16. Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2
17. Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1–15. https://doi.org/10.1080/14703297.2023.2195846
18. Fijačko, N., Gosak, L., Štiglic, G., Picard, C. T., & John Douma, M. (2023). Can ChatGPT pass the life support exams without entering the American heart association course? *Resuscitation*, *185*, 109732. https://doi.org/10.1016/j.resuscitation.2023.109732
19. Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487
20. Hsu, H.-Y., Hsu, K.-C., Hou, S.-Y., Wu, C.-L., Hsieh, Y.-W., & Cheng, Y.-D. (2023). Examining Real-World Medication Consultations and Drug-Herb Interactions: ChatGPT Performance Evaluation. *JMIR Medical Education*, *9*, e48433. https://doi.org/10.2196/48433
21. Ishak, N., Khairuddin, F. N., & Aziz, N. S. (2021). Perceived Postgraduate Research Supervisory Practices and Satisfaction towards Supervision among Supervisees at Faculty of Education. *IIUM Journal of Educational Studies*, *9*(2), 109–122. https://doi.org/10.31436/ijes.v9i2.278
22. Jacobsen, L. J., & Weber, K. E. (2023). *The Promises and Pitfalls of ChatGPT as a Feedback Provider in Higher Education: An Exploratory Study of Prompt Engineering and the Quality of AI-Driven Feedback* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/cr257
23. Jang, M. E., & Lukasiewicz, T. (2023). *Consistency Analysis of ChatGPT* (arXiv:2303.06273). arXiv. http://arxiv.org/abs/2303.06273

DOI: 10.5281/zenodo.14862530

24. Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115
25. Jordan, AlZu'bi, S., Mughaid, A., Jordan, Quiam, F., Jordan, Hendawi, S., & Jordan. (2023). Exploring the Capabilities and Limitations of ChatGPT and Alternative Big Language Models. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA3202820
26. Khuram, W., Wang, Y., Ali, M., Khalid, A., & Han, H. (2023). Impact of Supportive Supervisor on Doctoral Students' Research Productivity: The Mediating Roles of Academic Engagement and Academic Psychological Capital. *SAGE Open*, *13*(3), 21582440231185554. https://doi.org/10.1177/21582440231185554
27. King, R. C., Samaan, J. S., Yeo, Y. H., Kunkel, D. C., Habib, A. A., & Ghashghaei, R. (2023). *A multidisciplinary assessment of ChatGPT's knowledge of amyloidosis* [Preprint]. Cardiovascular Medicine. https://doi.org/10.1101/2023.07.17.23292780
28. Koubaa, A., Boulila, W., Ghouti, L., Alzahem, A., & Latif, S. (2023). *Exploring ChatGPT Capabilities and Limitations: A Critical Review of the NLP Game Changer* [Preprint]. MATHEMATICS & COMPUTER SCIENCE. https://doi.org/10.20944/preprints202303.0438.v1
29. Kushwah, L., & Navrouzoglou, P. (2022). Enhanced student satisfaction through effective supervisor-supervisee allocation: A case study. *Journal of Perspectives in Applied Academic Practice*, *10*(1). https://doi.org/10.56433/jpaap.v10i1.495
30. Li, S. W., Kemp, M. W., Logan, S. J. S., Dimri, P. S., Singh, N., Mattar, C. N. Z., Dashraath, P., Ramlal, H., Mahyuddin, A. P., Kanayan, S., Carter, S. W. D., Thain, S. P. T., Fee, E. L., Illanes, S. E., Choolani, M. A., Rauff, M., Biswas, A., Low, J. J. H., Ng, J. S., … Lim, M. Y. (2023). ChatGPT outscored human candidates in a virtual objective structured clinical examination in obstetrics and gynecology. *American Journal of Obstetrics and Gynecology*, *229*(2), 172.e1-172.e12. https://doi.org/10.1016/j.ajog.2023.04.020
31. Merriam-Webster. (n.d.). Feedback. *Merriam-Webster.Com Dictionary*. Retrieved November 27, 2023, from https://www.merriam-webster.com/dictionary/feedback
32. Mory, E., H. (2004). FEEDBACK RESEARCH REVISITED. In *Handbook of research on educational communications and technology* (pp. 745–783). Lawrence Erlbaum Associates Publishers.
33. Pereira, D., Flores, M. A., Simão, A. M. V., & Barros, A. (2016). Effectiveness and relevance of feedback in Higher Education: A study of undergraduate students. *Studies in Educational Evaluation*, *49*, 7–14. https://doi.org/10.1016/j.stueduc.2016.03.004
34. Phelps, D. L. (2013). *Supervisee Experiences of Corrective Feedback in Clinical Supervision: A Consensual Qualitative Research Study*. Marquette University.
35. Rao, A., Pang, M., Kim, J., Kamineni, M., Lie, W., Prasad, A. K., Landman, A., Dreyer, K., & Succi, M. D. (2023). Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. *Journal of Medical Internet Research*, *25*, e48659. https://doi.org/10.2196/48659
36. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003
37. Seeber, M., & Horta, H. (2021). No road is long with good company. What factors affect Ph.D. student's satisfaction with their supervisor? *Higher Education Evaluation and Development*, *15*(1), 2–18. https://doi.org/10.1108/HEED-10-2020-0044
38. Shihab, S. R., Sultana, N., & Samad, A. (2023). *Revisiting the use of ChatGPT in Business and Educational Fields: Possibilities and Challenges*. *02*(03).
39. Windon, S. R. (2020). The Relationship between Satisfaction with Supervisor and Demographic Variables among Extension Program Assistants. *Journal of Human Sciences and Extension*. https://doi.org/10.54718/XQGQ5060

DOI: 10.5281/zenodo.14862530