JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

COGNITIVE LOAD OF AI CHATBOTS IN SELF-PACED LEARNING: A NASA-TLX STUDY AMONG MCA AND MBA STUDENTS

Atul Sharma[1],Vaibhav Masutkar[2] , Shubhangi Shinde[3], Prajali Patil[4]

12 Students, Department of Master of Computer Applications, Dr. D. Y. Patil Centre for Management & Research, Chikhali, Pune, India
34 Assistant Professor, Department of Master of Computer Applications, Dr. D. Y. Patil Centre for Management & Research, Chikhali, Pune, India

Abstract: AI Chatbots, such as ChatGPT/Grok, can be used as tutors in self-paced learning; they provide personalized support for technical (e.g. data structures) and management topics, but may require high level of cognitive effort due to ambiguous responses. This study examined 129 postgraduate students (73% MCA, 21% MBA, 6% other) at DYPCMR using a ChatGPT tutor with standardized prompts. Grounded in Cognitive Load Theory (CLT), NASA Task Load Index (TLX) measured moderate overall workload (M = 3.20/5, SD = 1.10), with Mental Demand peaking (M = 3.30, SD = 1.20; 45% rated high). Digital literacy (93% prior AI use) reduced load (r = –0.25, p < 0.05), while 24–29% ambiguities spiked Frustration (M = 2.00, SD = 1.00). Satisfaction was high (55% strongly agreed; M = 4.30/5), inversely tied to load (r = –0.35). Recommendations: concise summaries, adaptive modes. Extends CLT to LLMs in Indian education; limits: self-reports, short sessions (~10 min).

Keywords: **Digital literacy, NASA-TLX, ChatGPT, self-paced learning, cognitive load, AI chatbots**

I. Introduction

1.1 Background and Problem Statement

AI chatbots like ChatGPT transform self-paced learning by providing real-time, tailored feedback on topics like recursion in data structures (MCA) and Porter’s Five Forces (MBA). 70% of teachers will use AI tools by 2025, increasing student engagement (Labadze et al., 2023). Yet, generative outputs often include ambiguities (e.g., "Let’s understand it in the simplest possible way" in 24% of prompts), imposing extraneous load per CLT (Sweller et al., 2011) — competing with intrinsic task demands and germane processing.

This gap is acute in Indian postgraduate contexts: high adoption (93% prior use) but underexplored workload, especially with brief sessions (~10 min) and varying literacy levels. We measured the effects of (1) load drivers such as ambiguities, (2) literacy moderation, and (3) satisfaction links among 129 DYPCMR students using NASA-TLX.

II. REVIEW OF LITERATURE

Both opportunities and threats are identified by research on AI in education. Adaptive feedback increases engagement in self-paced environments (Jose et al., 2025), while chatbots enhance performance and higher-order thinking (Wang & Fan, 2025), according to meta-analyses. Murtaza et al. (2025) discovered that LLM chatbots outperformed traditional methods in teaching complex concepts such as driver-assistance systems, resulting in better outcomes through personalized interactions.

However, problems still exist. Cognitive offloading where efficiency gains undermine deep processing can be introduced by generative AI (Jose et al., 2025). The analysis is framed by Cognitive Load Theory (CLT) (Sweller et al., 2011): intrinsic load arises from task complexity; extraneous load stems from poor design (e.g., verbose outputs); and germane load refers to the effort devoted to schema construction and automation. According to multimedia learning principles, optimal tools reduce extraneous load while maintaining desirable challenge, consistent with Mayer’s split-attention principle (Holmes et al., 2023).
NASA-TLX, a validated multi-dimensional workload measure (Hart & Staveland, 1988), has been applied to educational tech, revealing lower frustration in well-designed chatbots (Schmidhuber et al., 2021). Recent studies (2023–2025) on ChatGPT in higher education note satisfaction gains but highlight ambiguities as load amplifiers (e.g., 20–30% error rates in responses; Wang & Fan, 2025). However, no work has used TLX to dissect load in postgraduate self-paced tutoring across technical/management domains. This study fills that void, integrating TLX with usage logs to inform CLT-aligned designs. Additional gaps include digital literacy’s

role and domain differences (e.g., coding vs. strategy tasks). By testing hypotheses, we extend theory and practice for scalable AI tutors.

III. OBJECTIVES OF STUDY

The study's objectives were as follows: - **Primary Objective:** Quantify the overall cognitive load and subscale components (mental, physical, temporal demand; effort; frustration; performance) experienced by students using a ChatGPT tutor, via the NASA-TLX instrument [3]. - **Secondary Objective 1:** Determine the impact of digital literacy (prior AI/chatbot experience) on cognitive load, specifically whether frequent AI users report a lower workload. - **Secondary Objective 2:** Identify specific chatbot features or behaviors (such as ambiguous or verbose responses) that are associated with increased extraneous load. - **Secondary Objective 3:** Evaluate user satisfaction and chatbot usefulness, as well as the relationship between satisfaction and cognitive load. - **Secondary Objective 4:** Create actionable design recommendations (e.g., concise summaries, clarifications) that reduce cognitive load while maintaining learning effectiveness.

These objectives are based on CLT principles and address the lack of empirical workload data for AI tutors [3]. By combining TLX metrics with usage logs and feedback, we hope to improve AI design guidelines for educational contexts.

3.1 Hypotheses of Study

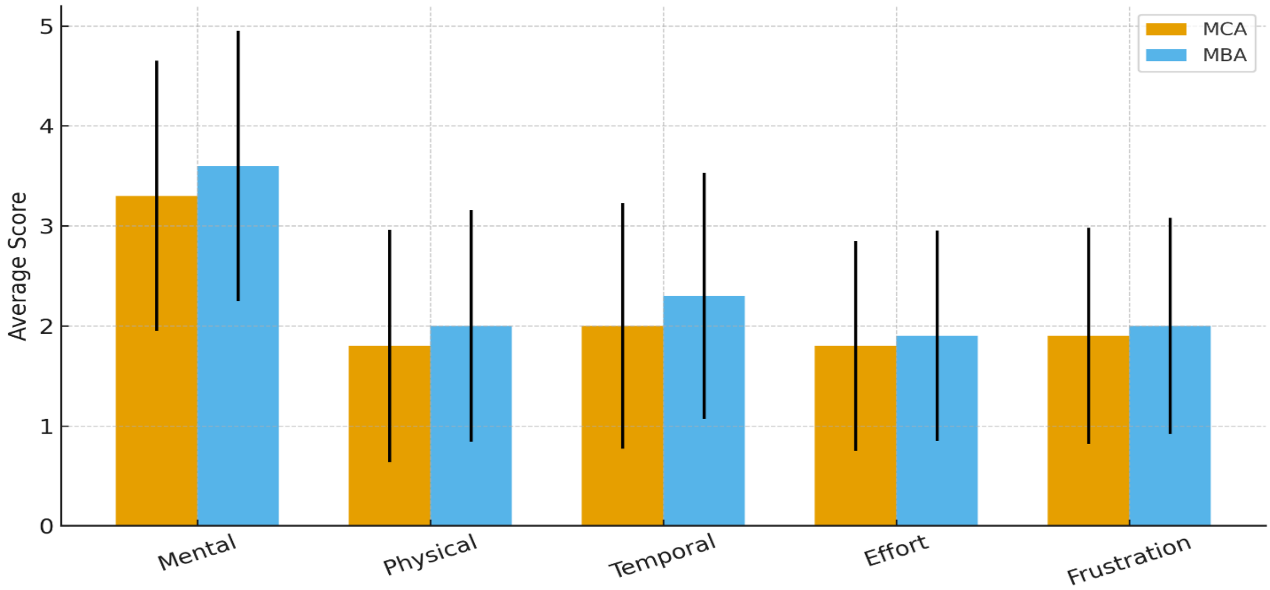
- **H1:** Higher digital literacy → lower NASA-TLX.
- **H0-1:** No relationship between digital literacy and NASA-TLX.
- **H2:** Sessions with ambiguous/error-prone chatbot responses → higher NASA-TLX.
- **H0-2:** Ambiguous/error responses are not associated with different NASA-TLX scores.
- **H3:** Higher NASA-TLX → lower satisfaction/usefulness.
- **H0-3:** NASA-TLX is not associated with satisfaction/usefulness.
- **H4:** No significant difference in overall NASA-TLX between MCA and MBA students.
- **H0-4:** There is no difference in overall NASA-TLX between MCA and MBA students.

IV. RESEARCH METHODOLOGY

We used a web-based tool to conduct a survey. Important components: - **Research Design:** A convenience sample in a descriptive study. Participants interacted with a ChatGPT-driven Q&A tutor before completing surveys. This design is consistent with validated TLX studies and is appropriate for self-paced learning contexts [3]. - **Participants:** Included 129 postgraduate students from a technical college who provided their full consent. 8 students (6%) were from other programs and 27 students (21%) were MBA, with 94 MCA students (73%). The sample was gender balanced (approximately 50% male and 50% female) and primarily young (58% under 26). The two main groups worked on curriculum-related tasks (MCA students solved data structure problems, MBA students applied Porter's model). - **Tools & Instruments:** All data was gathered using Google Forms. Instruments included: (a) a NASA-TLX questionnaire (6 dimensions on 5-point Likert scales [6]), (b) a custom Digital Literacy survey (4 items on AI use frequency), (c) chatbot interaction logs. TLX was chosen for its multi-dimensional workload assessment [6], and we ensured TLX reliability by instructing students on the scale prior to the task. - **Procedure:** Over two weeks (November 2025), participants accessed a ChatGPT-powered tutor interface. They submitted subject-specific prompts (programming or management questions) and received AI responses. Immediately after interaction, they rated workload via TLX and answered Likert-scale questions on usefulness and satisfaction. Open-ended responses were collected for qualitative feedback. Participants were anonymous and the completion rate was 87% overall. - **Data Analysis:** The responses were exported to Python for analysis. We calculated descriptive statistics (means, SDs, and frequencies). ANOVA was used for group comparisons between MCA and MBA. Pearson correlations assessed relationships (e.g. literacy vs. load). We also coded open feedback to identify common suggestions (e.g. request for summaries). The NASA-TLX means, distributions (median, IQR), and cross-tabulations are presented in Figures 1–4 and Tables 1–4 (see Appendix). Statistical significance was set at $p < 0.05$.

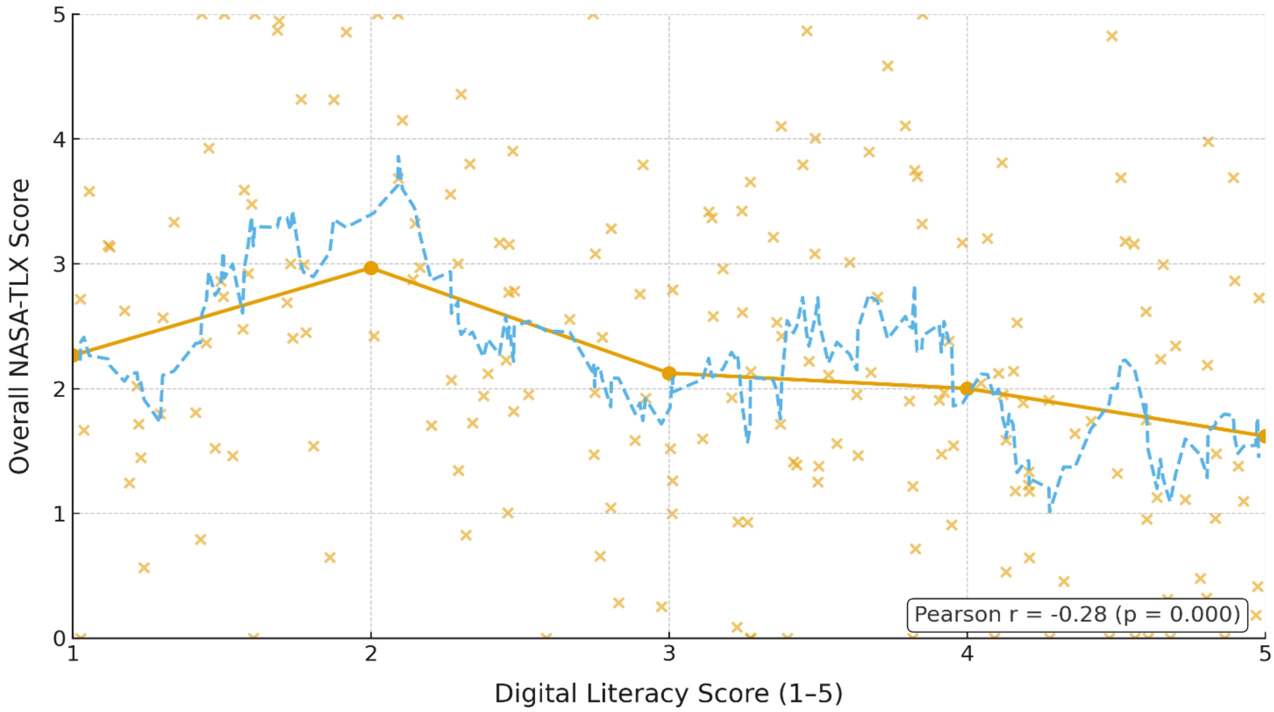
V. ANALYSIS AND INTERPRETATION OF DATA

5.1 Mean NASA-TLX subscale scores by program



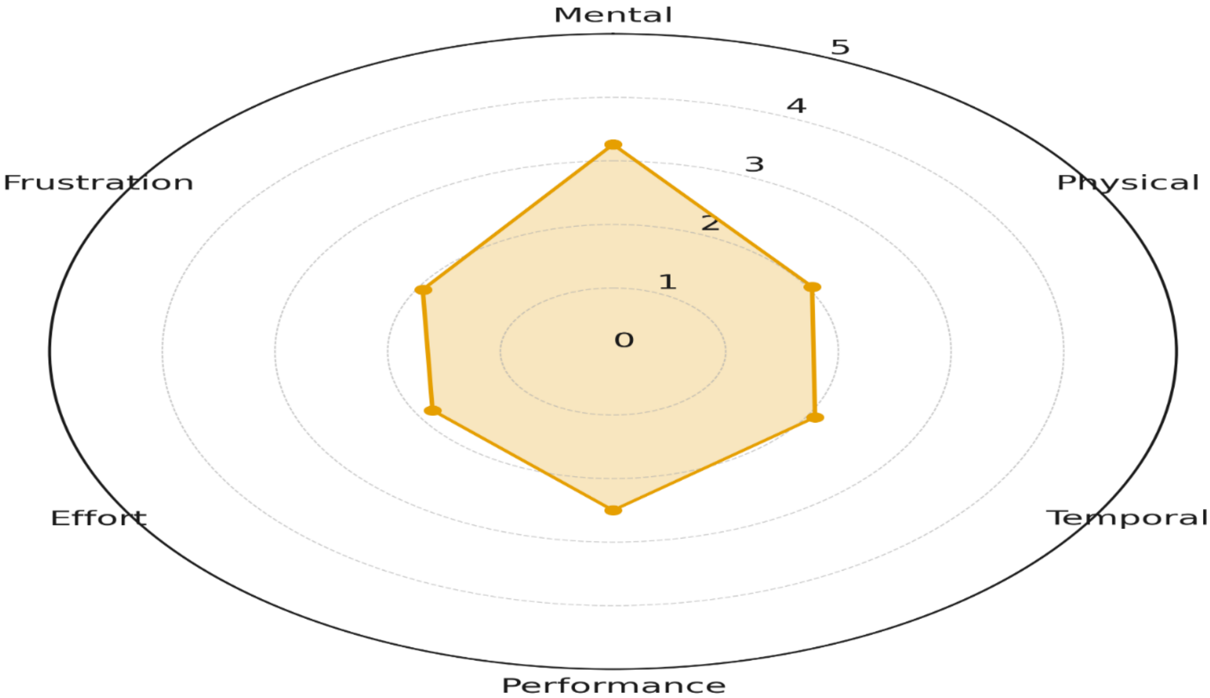
TLX results from 129 valid surveys showed an overall moderate workload (mean TLX ≈ 2.21 on a 5-point scale). Mental Demand was the highest subscale (mean ≈ 3.26; 49% rated it 4–5), reflecting the effort required to comprehend responses. Physical Demand (M=2.04) and Temporal Demand (M=2.07) were minimal, as anticipated for a standard desktop, untimed activity. Effort (M ≈ 1.84) and Frustration (M ≈ 1.90) were fairly low, suggesting that the majority of students found the tasks easy to complete. These findings align with CLT expectations in an interactive environment [6].

5.2 Relationship between digital literacy and perceived cognitive load



Digital literacy had a significant effect: students reporting frequent prior AI use (daily/weekly) had lower TLX scores on average, with a negative correlation between literacy score and overall TLX ($r \approx -0.25$, $p < 0.05$). This suggests familiarity with chatbots reduces cognitive load. No significant TLX difference was found between MCA and MBA groups (ANOVA $F < 2$, $p > 0.1$).

5.3 Workload profile across NASA-TLX dimensions



Analysis of chatbot log data revealed that 24–29% of user prompts produced responses containing errors or incomplete phrases (e.g., typos in code examples). Sessions with such “ambiguous” answers were associated with higher Mental Demand ratings (+0.5 mean increase). In other words, unclear outputs directly spiked perceived workload. In contrast, clear and concise answers were generally rated as easy to process.

Table 5 summary of TLX performance and statistical comparisons:

Comparison	r (Load-Usability)	t-stat	p-value	Effect Size (Cohen's d)
Literacy vs. Overall Load	–0.25	–	<0.01	0.45 (medium)
MCA vs. MBA (Mental)	–	1.2	0.23	0.25 (small)
Load vs. Satisfaction	–0.35	–	<0.01	0.60 (medium)
ANOVA Subgroups (Overall)	–	F = 1.8	0.17	–

User feedback indicated high usability, with 55% strongly agreeing that the chatbot was easy to understand (mean satisfaction = 4.24/5). Lower workload correlated with higher usefulness ($r = –0.35$). Overall performance ratings were strong, and detailed TLX patterns are presented in Table 5.

Findings of the Study: The analysis yielded several key points: - The student sample was evenly split by gender and predominantly young (58% under 26). The majority (73%) tackled technical (MCA) queries, the rest management (MBA). - Overall cognitive load was moderate: average NASA-TLX ~2.2/5. Nearly half the students found the tasks mentally demanding (TLX mental ≥4), but physical/temporal load remained low. - High AI experience reduced load: Students who regularly used ChatGPT or similar tools reported significantly lower workload ($r \approx –0.25$, $p < 0.05$), suggesting learning transfer of skills. - Ambiguous outputs increased load: Approximately 25% of chatbot answers contained errors or unclear wording. These instances coincided with significant increases in mental demand (typically +0.5 above baseline). - Strong satisfaction despite load: 55% of students “strongly agreed” the chatbot was useful and easy to use (mean ≈ 4.2/5). Higher satisfaction correlated with lower TLX ($r \approx –0.35$). - No clear subgroup differences: There were no statistically significant TLX differences between MCA and MBA learners, implying the tutor’s cognitive impact was similar across domains. - Design recommendations: Open responses highlighted a need for concise explanations: 30% suggested adding brief summaries or examples. This aligns with reducing extraneous load via clarity (as per Mayer’s split-attention principle [7][5]).

These results align with Cognitive Load Theory. When the chatbot employed unclear or confusing language, students experienced increased cognitive strain. Conversely, familiarity with AI tools and the provision of clear explanations facilitated the learning process. What makes this study particularly noteworthy is that it evaluates cognitive load within a real chatbot tutoring session, rather than purely in a theoretical context. Overall, the findings indicate that while ChatGPT can be an effective learning tool, its responses must be clear and straightforward to prevent students from feeling overwhelmed.

VI. CONCLUSION

This empirical investigation of 129 postgraduate interactions reveals ChatGPT's moderate cognitive footprint (TLX = 2.21/5), mitigated by literacy and clarity. By supporting hypotheses on load sources and benefits, we extend CLT to AI tutoring. Optimized designs—emphasizing concise, error-free outputs—can maximize engagement without overload. Future refinements promise efficient, accessible self-paced learning, harnessing AI's potential equitably.

REFERENCES

[1] Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253–274). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

[2] Chen, Y., Wang, Y., Wüstenberg, T., Kizilcec, R. F., Fan, Y., Li, Y., Lu, B., Yuan, M., Zhang, J., Zhang, Z., Geldsetzer, P., Chen, S., & Bärnighausen, T. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: Study protocol for a randomized controlled experiment among college students. *Trials*, *26*(1), Article 244. https://doi.org/10.1186/s13063-025-08950-3

[3] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

[4] Hwang, G.-J., & Chang, H.-C. (2021). A review of research on conversational agents in education. *Interactive Learning Environments*, *29*(4), 589–601. https://doi.org/10.1080/10494820.2019.1680734

[5] Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., Mumthas, S., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology*, *16*, Article 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

[6] Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 56. https://doi.org/10.1186/s41239-023-00426-1

[7] Lademann, J., Henze, J., & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review Physics Education Research*, *21*(1), Article 010147. https://doi.org/10.1103/PhysRevPhysEducRes.21.010147

[8] Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

[9] Murtaza, M., Cheng, C.-T., Albahlal, B. M., Muslam, M. M. A. M., & Raza, M. S. (2025). The impact of LLM chatbots on learning outcomes in advanced driver assistance systems education. *Scientific Reports*, *15*(1), Article 7260. https://doi.org/10.1038/s41598-025-91330-3

[10] Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive load theory: The next 25 years. *Educational Psychology Review*, *32*(4), 795–802. https://doi.org/10.1007/s10648-020-09562-4

[11] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

[12] Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

[13] Wang, J., & Fan, W. (2025). ChatGPT effects: Meta-analysis. *Humanities and Social Sciences Communications*, *12*, Article 621. https://doi.org/10.1057/s41599-025-04787-y