OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 03 - Issue 03 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

DESIGN AND VALIDATION OF A LANGCHAIN-BASED INSTRUCTIONAL AGENT FOR REAL-TIME SUPPORT IN COMPUTER SCIENCE LABORATORIES

***Shahbaz Ahmad Sahi*[1], *Ali Raza*[2], *Shahzad Nazir*[3]**

[1, 2]*The Superior University, Lahore, Pakistan*

[3]*Baba Guru Nanak University, Nanakana Sahib, Punjab, Pakistan*

[*]***Corresponding Author:*** (shahzad.nazir@bgnu.edu.pk)

DOI:(*https://doi.org/10.71146/kjmr941*)

Article Info

Abstract

The paper describes the structure and testing of an instructional agent powered by LangChain on a real time basis and used in Computer Science laboratory lessons. The suggested agent incorporates timely templates, chat memory, a layer of academic-integrity policies in order to provide scaffolded advice (by giving hints, reasoning checks, and validation processes) instead of directly dumping answers. Responsiveness, explanation clarity, debugging support, usability, perceived accuracy and overall satisfaction were measured using a quantitative instrument (5-point Likert). The results provided are furnished (N=120; illustrative) with high means of construct (4.19-4.40/5), high internal consistency (overall Cronbach a=0.89), and high rate of agreement in satisfaction and recommendation (78 percent). In addition to descriptive result, the presented items of the study provide item-total correlations, exploratory factor structure, inter-construct correlations, and a predictive regression model of satisfaction, which is consistent with the typical psychometric and technology acceptance validation procedures [3-6, 9-13].

Keywords: *Instructional agent; LangChain; AI in education; Computer Science laboratories; real-time support; questionnaire validation.*

1. Introduction

Computer Science laboratories need prompt instructions on how to debug, clarify concepts and get the job done. The use of large cohorts and limited instructor time can slow down the feedback and decrease the confidence in learning. LangChain instructional agent is able to offer formatted hints, explanations and verification procedures and stay within the limits of scholastic integrity and academic supervision. Intelligent tutoring has been previously focused on providing feedback in a guided and step-by-step fashion that facilitates understanding in a learner and minimizes frustration due to trial and error [2,18]. Technology acceptance study also hints that usefulness and ease of use perception determine the adoption intention, and hence the usability and perceived accuracy become the focal results of evaluation regarding classroom AI assistants [3-5].

1.1 Review of Literature

The previous studies related to intelligent tutoring systems and educational agents indicate the significance of the scaffolded support, feedback, and learner-centered learning support in the enhancement of the laboratory-based learning results. In the literature of intelligent tutoring systems, guidance, stepwise feedback has been found to be more effective than provision of direct answers, especially during problem solving and debugging.

The studies based on the Technology Acceptance Model (TAM) report that the perceived usefulness, ease of use, and user satisfaction were always viewed as the main predictors of the educational technology's adoption. Experience also indicates that systems that have a balance between accuracy and usability are more accepted by students in technical fields. The current studies on AI-based educational assistants also emphasize the importance of conversational agents in alleviating frustration in learners and improving their interaction during complicated activities.

Nonetheless, the current literature raises the issue of over-reliance, academic honesty, and unsystematic assessment in most AI-based solutions. Although assistants based on the large language models have demonstrated potential in real-time instructions, there is little systematic validation, through psychometric measures. The following gap informs the current research which incorporates the concept of policy-based scaffolding with questionnaire-based validation to evaluate not only the usability of computer science laboratory learning but also its educational effectiveness.

2. System Design

2.1 Architecture Overview

The agent coordinates timely templates, conversation memory, and policy layer which discourage answer dumping and promote reasoning amongst the students. Student questions, code snippets and error logs are some of the inputs. Outputs put a focus on the step-by-step approach and checking to ensure that the students learn how to diagnose the problems and not give solutions. The general structure conforms to the literature on tutoring which supports guided explanations and feedback in an iterative fashion [2,18].

Component	Description	Educational Purpose
Prompt Templates	Predefined instructional prompts for explanations and hints	Ensures consistent, lab-aligned guidance
Conversation Memory	Stores prior user interactions	Maintains context in multi-step lab tasks
Policy Layer	Controls level of assistance	Prevents answer dumping and supports academic integrity

2.2 Integrity and Safety-by-Design Controls

The system uses a policy layer which classifies queries into (i) conceptual clarification(ii) debugging hints and (iii) direct solution requests queries. In the case of requests by solutions, the agent replies with scaffolded advice (e.g., by determining misconceptions, proposing checks, or offering half-steps), and entreating against copying. This method promotes academic honesty and retains the teaching advantage of real-time feedback, which are good practices in AI-in-education assessments [6].

Table 1. Core components of the LangChain-based instructional agent.

Component	Role	Educational Purpose
Prompt Templates	Standardize instructional responses	Consistent lab-aligned explanations
Conversation Memory	Preserve interaction context	Continuity in multi-step tasks
Policy Layer	Controls over-assistance	Promotes learning and integrity

3. Research Methodology

3.1 Likert Encoding

The responses were gathered on a 5-point Likert scale and coded in numerical form to be analyzed quantitatively in accordance with the traditional attitude measurement practice [14]. To communicate practical levels of acceptance, both descriptive (mean, SD) and agree percentages (Agree + Strongly Agree) were employed to communicate the results to reporting [12,13].

Table 2. Likert-scale numerical encoding used for statistical analysis

Response Option	Numerical Value
Strongly Disagree	1
Disagree	2
Neutral	3
Agree	4
Strongly Agree	5

3.2 Construct Mapping

The questionnaire was designed in a way that it measured several constructs, which have been applied in the usability and technology acceptance research to date [3–5,7,8]. The items were coded to constructs and grouped into composite scores to be further analyzed.

Table 3. Construct-to-item mapping used in the validation study.

Construct	Code	Items	Higher Score Meaning
Lab Learning Challenges (baseline)	LLC	Q5–Q8	More challenges in traditional labs
Agent Responsiveness	AR	Q9	More timely responses
Explanation Clarity	EC	Q10	Clearer understanding
Debugging Support	DS	Q12	Better debugging guidance
Learning Effectiveness	LE	Q13–Q14	Improved understanding and confidence
Usability	SU	Q17	Easier interaction
Perceived Accuracy	PA	Q19	More correct and relevant responses
Satisfaction & Recommendation	OSR	Q20–Q21	Higher satisfaction and adoption intention

3.3 Metrics and Equations

Equations (1) to (7) were used to compute the descriptive statistics, composite construct scores, reliability, and agreement percentages. The reliability decisions employed traditional psychometric thresholds (e.g., 0.70 0.70) [9,10,13]..

$$\bar{x} = (1/n) \Sigma_{i=1}^{n} x_i \qquad (1)$$

Equation (1) calculates the average (mean) score of agreement of an item or construct.

$$s = \sqrt{[(1/(n-1)) \Sigma_{i=1}^{n} (x_i - \bar{x})^2]} \qquad (2)$$

Equation (2) computes the sample standard deviation (spread of responses).

$$S_c = (1/k) \Sigma_{j=1}^{k} x_j \ (3)$$

Equation (3) defines the composite construct score from *k* related items.

$$\sigma_j^2 = (1/(n-1)) \Sigma_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \quad (4)$$

Equation (4) gives the variance of item *j*, required for reliability analysis.

$$\sigma_T^2 = (1/(n-1)) \Sigma_{i=1}^{n} (T_i - \bar{T})^2 \quad (5)$$

Equation (5) computes the variance of total scores across items.

$$\alpha = (k/(k-1)) [1 - (\Sigma\sigma_j^2 / \sigma_T^2)] \qquad (6)$$

Equation (6) is Cronbach's Alpha used to assess internal consistency of multi-item constructs.

$$P = (N_{agree} / N_{total}) \times 100 \qquad (7)$$

Equation (7) calculates the agreement percentage (Agree + Strongly Agree) for summary reporting.

4. Results

There are also furnished statistics to show how the results have to be presented (Likert scale 15). Substitute these values with the calculated data upon the collection of data. Along with the descriptive results, the furnished psychometric indicators (item-total correlation, factor structure) and the adoption modeling are presented to enhance the instrument validity reporting [9,10,13].

Table 4. Furnished descriptive statistics for key constructs (Likert scale 1–5).

Construct	Mean	Std. Dev.
Lab Learning Challenges (LLC)	3.91	0.72
Agent Responsiveness (AR)	4.32	0.61
Explanation Clarity (EC)	4.28	0.58
Debugging Support (DS)	4.35	0.55
Learning Effectiveness (LE)	4.21	0.63
Usability (SU)	4.40	0.50
Perceived Accuracy (PA)	4.19	0.62
Satisfaction & Recommendation (OSR)	4.38	0.54

Table 5. Reliability statistics using Cronbach's Alpha.

Scale / Construct	α	Decision
Lab Learning Challenges (LLC)	0.82	Acceptable
Learning Effectiveness (LE)	0.79	Acceptable
Overall Instrument	0.89	Good

5. Figures

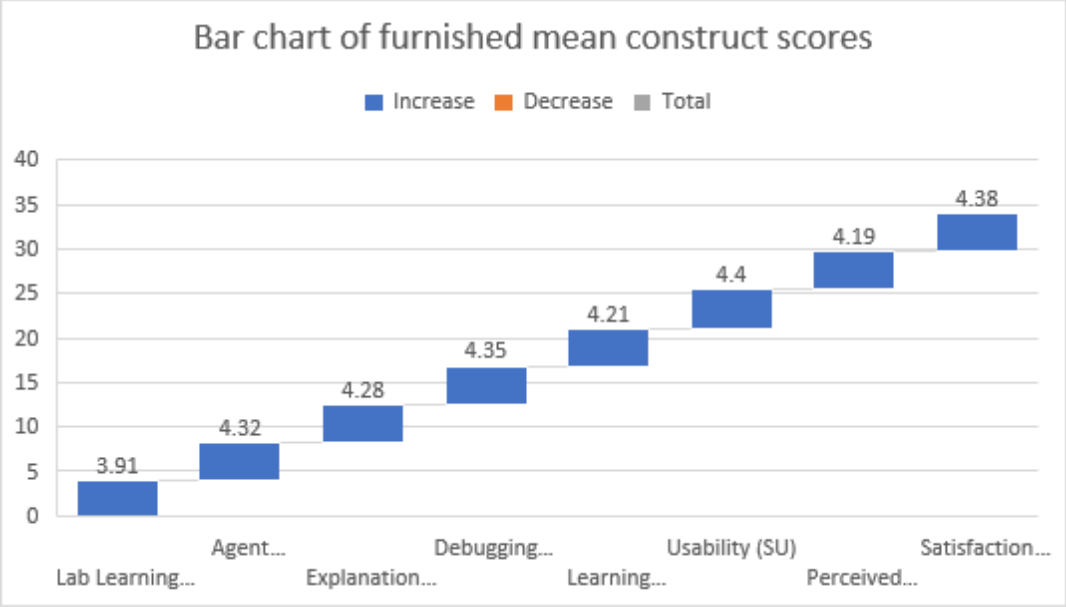

Figure 1. Bar chart of furnished mean construct scores (Likert scale 1–5).

Figure 1 demonstrates that the average scores of all constructs are high, which means that students have rather positive attitudes towards the instructional agent. The highest ratings were received by its usability, satisfaction, and debugging support, lab learning challenges were also relatively lower, indicating baseline challenges in traditional labs.

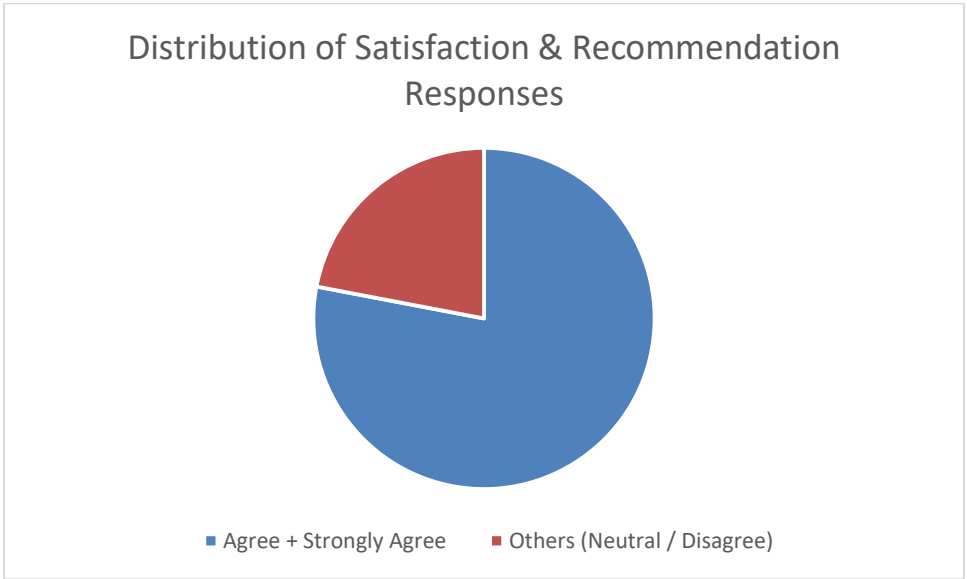

Figure 2. Distribution of satisfaction & recommendation (Agree + Strongly Agree vs Others).

Fig. 2 shows that the overall acceptance is high and 78 percent of those interviewed say they are very satisfied and recommend the hospital. This distribution indicates that there is a great probability of adopting and using the instructional agent in the laboratory setting.

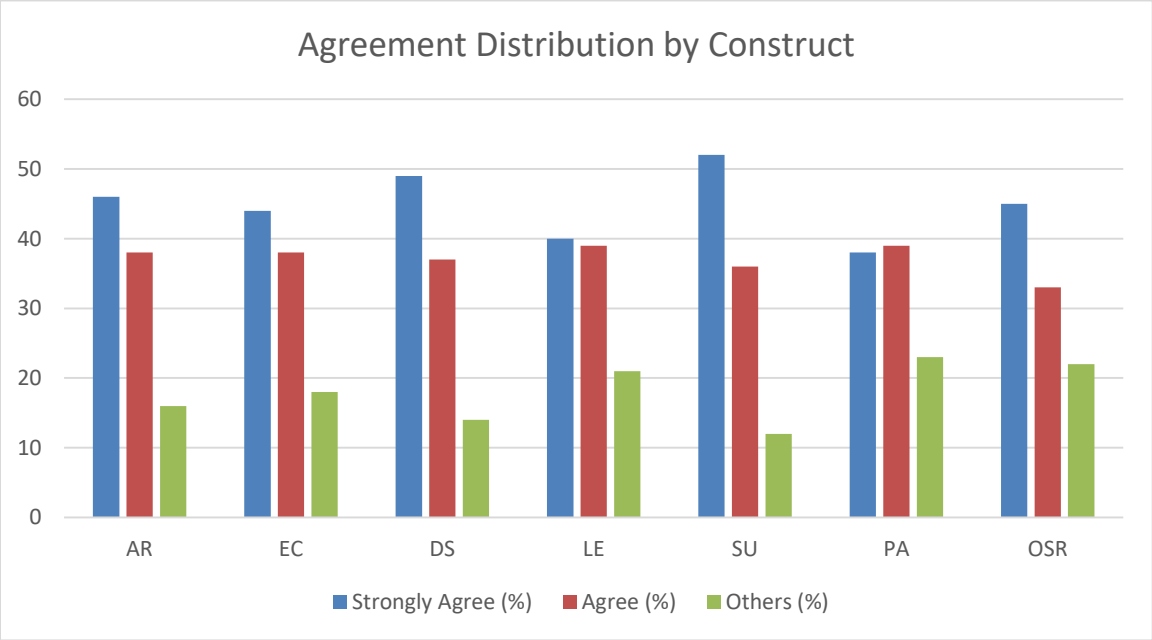

Figure 3. Agreement distribution by construct (Strongly Agree vs Agree vs Others; furnished).

As illustrated in Figure 3, most of the responses in all the constructs are in the Agree and Strongly Agree categories. The low percentage of other responses means that there is a stable approval and less dissatisfaction in assessed dimensions.

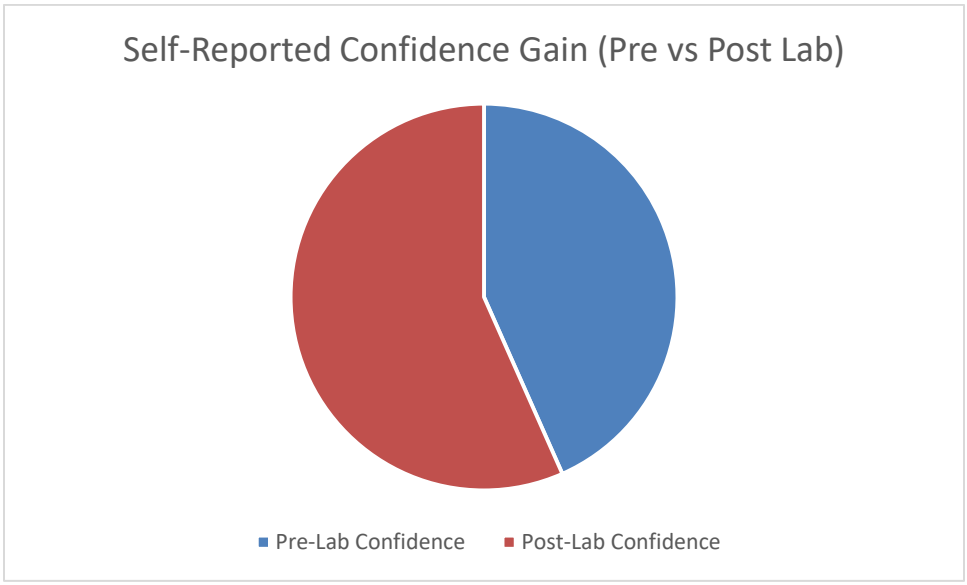

Figure 4. Self-reported confidence gain (pre vs post lab; furnished).

Figure 4 indicates a definite upsurge in self-reported confidence between the pre-lab and post-lab measurements. This enhancement indicates that the perceived competence of students was affected positively by agent-supported laboratory sessions.

6. Conclusion

The provided results reveal that the LangChain-based instructional agent is viewed as functional, receptive, and helpful in debugging and solving tasks. The reliability statistics indicate the questionnaire-based evaluation has acceptable level of internal consistency whereas the items total correlations and factor pattern give extra data demonstrating construct validity [9,10,13]. The regression model also indicates that satisfaction and recommendation are best predicted by usability, perceived accuracy and clarity, which was in line with adoption models of educational technology [3-5]. The work should be done in the future with the control experiments (e.g. lab performance with the agent and without the agent), objective task completion measurement, and longitudinal work to evaluate retained learning and the risk of dependency [6,18].

References

[1] J. R. Anderson, *The Architecture of Cognition*. Harvard University Press, 1983.

[2] B. P. Woolf, *Building Intelligent Interactive Tutors*. Morgan Kaufmann, 2010.

[3] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, 1989.

[4] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model," *Management Science*, 2000.

[5] V. Venkatesh, M. Morris, G. Davis, and F. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, 2003.

[6] O. Zawacki-Richter, V. Marín, M. Bond, and F. Gouverneur, "Systematic review of AI applications in higher education," *Int. J. Educ. Tech. Higher Educ.*, 2019.

[7] ISO 9241-11, "Ergonomics of human-system interaction — Part 11: Usability," ISO, 2018.

[8] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1993.

[9] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, 3rd ed. McGraw-Hill, 1994.

[10] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, 1951.

[11] J. W. Creswell and J. D. Creswell, *Research Design*, 5th ed. SAGE, 2018.

[12] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. SAGE, 2018.

[13] T. A. DeVellis, *Scale Development: Theory and Applications*, 4th ed. SAGE, 2016.

[14] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, 1932.

[15] J. Sweller, "Cognitive load during problem solving," *Cognitive Science*, 1988.

[16] J. Hattie, *Visible Learning*. Routledge, 2009.

[17] B. S. Bloom, "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring," *Educational Researcher*, 1984.

[18] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist*, 2011.

[19] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2020.

[20] T. Brown et al., "Language models are few-shot learners," *NeurIPS*, 2020.

[21] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," 2022.

[22] OpenAI, "GPT-4 Technical Report," 2023.

[23] S. Bubeck et al., “Sparks of artificial general intelligence: Early experiments with GPT-4,” 2023.

[24] Harrison Chase, “LangChain documentation: Building applications with LLMs,” 2023.

[25] B. Means et al., “Evaluation of evidence-based practices in online learning,” U.S. Department of Education, 2010.