

Developing Students' Practical Programming Competencies Through AmaliyLab: An Adaptive Digital Learning Approach and Multi-Institutional Evaluation



Sulaymonov Zulqaynar Rajabali o'g'li

Jizzakh Branch of the National University of Uzbekistan, Uzbekistan

Received: 12 February 2026; **Accepted:** 08 March 2026; **Published:** 31 March 2026

Abstract: This article examines the development of students' practical programming competencies through AmaliyLab, an adaptive software-supported pedagogical approach developed within research on improving software for students' practical activities. Unlike uniform practical instruction, the proposed approach links initial diagnostics, differentiated tasks, performance monitoring, graduated feedback, reflection and subsequent task selection. The research draws on a quasi-experimental study associated with Programming fundamentals during the 2024–2025 and 2025–2026 academic years at Fergana State University, Gulistan State University and Karshi State University. The aggregated dataset comprises 307 students, including 153 in experimental and 154 in control groups. The combined proportion of students classified at high or medium competency levels increased from 41.18% to 77.78% in the experimental groups and from 40.91% to 62.99% in the control groups. Pearson's chi-square test revealed no baseline difference in the three-level distributions ($\chi^2 = 0.0023$, $df = 2$, $p \approx 0.99887$) and a statistically significant difference at the final assessment ($\chi^2 = 23.4514$, $df = 2$, $p \approx 0.0000081$; Cramér's $V \approx 0.2764$). These findings support a favorable association between the AmaliyLab-supported teaching approach and competency-level outcomes, while the aggregated and quasi-experimental nature of the data limits causal attribution.

Keywords: AmaliyLab, adaptive digital learning, practical programming competencies, differentiated tasks, formative feedback, pedagogical experiment, learning analytics, Pearson chi-square.

Introduction: The development of practical competencies is an essential outcome of contemporary higher education. In programming instruction, practical activity extends beyond reproducing source code: students must analyze problem requirements, design algorithms, implement and test solutions, identify errors, explain results and work independently. Digital learning tools can support these activities, but uniform task sequences may not accommodate variation in students' prior preparation or their need for instructional support [1–4].

The dissertation underlying this article, "Improving software for developing students' practical activities," focuses on the teaching of Programming fundamentals.

Its central software product, AmaliyLab, is conceptualized as a pedagogical management instrument within the "student–AmaliyLab–teacher" relationship, rather than as a replacement for the teacher. The primary innovation is the joint adaptation of task complexity and the amount of pedagogical assistance in response to diagnosed needs and observed task performance.

Accordingly, the study addresses three questions: (RQ1) how can the AmaliyLab-supported instructional process be structured to develop practical programming activity; (RQ2) what changes are observed in competency-level distributions over the intervention; and (RQ3) do the experimental and

control groups differ statistically at baseline, during implementation and at the final assessment?

METHODS

Pedagogical and software-supported approach. The AmaliyLab instructional model follows a recurring cycle: initial diagnostics → learner profile → practical task matched to readiness → observation of solution processes → assessment of results and errors → differentiated feedback → student reflection → selection of subsequent tasks. Its methodological architecture includes diagnostic, task-selection, performance-monitoring, feedback, analytical and teacher-guidance functions. These functions describe the dissertation’s pedagogical model; this paper does not independently verify any specific LMS protocol, automated grading implementation, machine-learning algorithm or application programming interface.

Practical assignments are differentiated into three complexity levels: L1, application of basic programming constructions; L2, combined use of several constructions in multi-step tasks; and L3, independent selection and comparison of algorithms, testing and improvement of solutions. For students who solve tasks independently, task difficulty may increase while

the level of assistance decreases. Repeated errors may trigger additional hints, alternative explanations or reduced task complexity. Teachers interpret the monitoring evidence and can intervene in individual or group learning.

Unlike the earlier manuscript, no claim is made here that AmaliyLab currently uses Bayesian Knowledge Tracing, fixed probabilities of learning, a six-node prerequisite graph, LTI 1.3 or xAPI: these implementation-specific claims are not established by the dissertation evidence used for this revision. The original synthetic-learner simulation is therefore not presented as empirical validation of AmaliyLab.

Study design and participants. A quasi-experimental comparison between experimental groups using the AmaliyLab-supported methodology and control groups receiving conventional practical instruction was reported in dissertation materials. The study period covers the 2024–2025 and 2025–2026 academic years, with practical instruction linked to Programming fundamentals. Three participating universities were Fergana State University, Gulistan State University and Karshi State University. The consolidated group counts are shown in Table 1.

Table 1. Participant distribution by institution

University	Experimental	Control	Total
Fergana State University	51	52	103
Gulistan State University	50	52	102
Karshi State University	52	50	102
Total	153	154	307

The dissertation data are summarized by institutions, groups and stages. The supplied materials do not document, in a verifiable individual-level dataset, the precise assignment procedure, participant-level matching, whether cohorts overlap across academic years, or the duration of exposure for each learner. These aspects require confirmation before stronger causal or longitudinal inferences can be made.

Indicators and competency levels. Practical activity was considered in four interconnected dimensions:

motivational (independent engagement with assignments); cognitive (problem analysis and algorithmic reasoning); practical-operational (coding, execution, testing and debugging); and reflective-evaluative (verification of outputs, explanation of errors and planning improvements). Aggregated outcomes were classified as high, medium or low. Identical category meanings were used for the group comparison; the original individual scoring rubric and numerical cutoffs should be supplied as supplementary

documentation.

Statistical analysis. For each assessment stage, the experimental and control frequency distributions across the three ordered competency levels were compared using Pearson's chi-square test for a 2×3 contingency table (degrees of freedom = 2). Cramér's V summarizes the strength of association. Proportions at high or medium level were computed as (high + medium)/group size $\times 100\%$, with changes expressed in percentage points. These are marginal group-level comparisons; they do not recover students' individual transitions. With no validated continuous individual

scores, post-test t statistics, Cohen's d, paired-test statistics, reliability coefficients or correlations with platform usage were not estimated.

RESULTS

Competency distributions across assessment stages.

Table 2 reproduces the consolidated three-level distributions reported in the dissertation. Initially, the experimental and control groups were closely aligned. During the formative stage and especially at the final assessment, the distributions diverged in favor of the experimental groups.

Table 2. Aggregated competency levels across the three stages (number of students)

Stage	Group	n	High	Medium	Low
Initial	Experimental	153	16	47	90
Initial	Control	154	16	47	91
Formative	Experimental	153	37	53	63
Formative	Control	154	19	61	74
Final	Experimental	153	58	61	34
Final	Control	154	22	75	57

Table 3. Pearson chi-square comparisons of group distributions

Stage	χ^2	df	p value	Cramér's V
Initial	0.0023	2	0.99887	0.0027
Formative	7.2271	2	0.02696	0.1534
Final	23.4514	2	0.0000081	0.2764

At baseline, the null hypothesis of equal competency-level distributions could not be rejected ($p \approx 0.99887$). The formative comparison was statistically significant ($p \approx 0.02696$), and the final group difference was stronger ($p \approx 0.0000081$). At the final stage, 58 of 153 students in the experimental group were at the high level (37.91%), compared with 22 of 154 (14.29%) in the control group. The corresponding low-level counts were 34 and 57.

Magnitude and institutional pattern of change. The proportion at high or medium competency level increased by 36.60 percentage points in the experimental groups (41.18% $\rightarrow$ 77.78%) and by 22.08 percentage points in the control groups (40.91% $\rightarrow$ 62.99%). The between-group difference at the final stage was 14.79 percentage points. Figure 1 displays the changes.

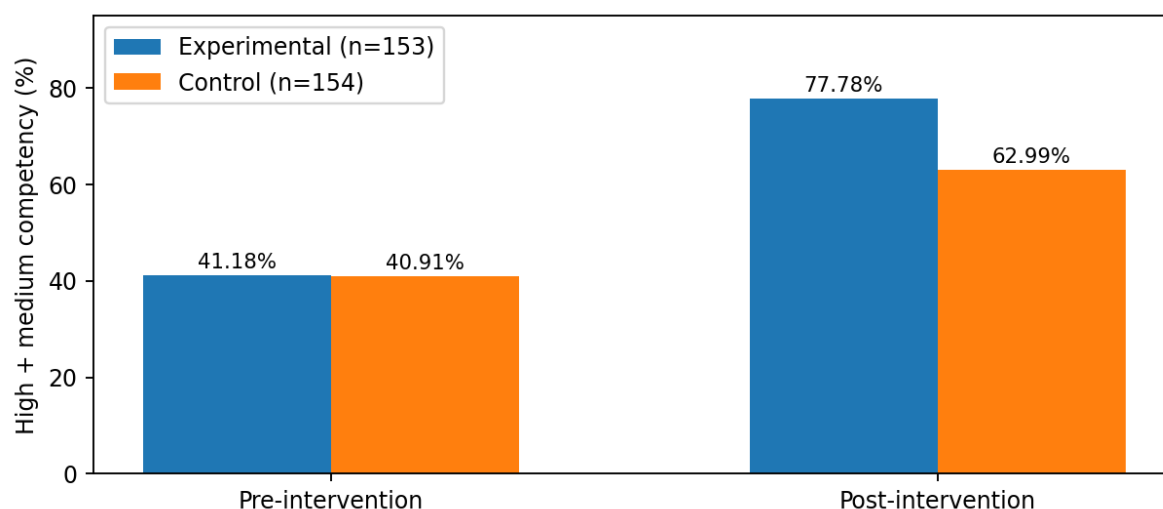


Figure 1. Combined high- and medium-level shares in the experimental and control groups.

Table 4. High- and medium-level shares by institution (%)

Institution	Experimental initial	Experimental final	Control initial	Control final
Fergana State University	41.18	70.59	40.38	51.92
Gulistan State University	40.00	80.00	42.31	61.54
Karshi State University	42.31	82.69	40.00	76.00
Overall	41.18	77.78	40.91	62.99

An institutional breakdown shows a positive shift at all three universities, although the final gap between groups differs by institution. For the binary high-or-medium outcome, the dissertation reports a final overall difference of 14.79 percentage points (approximate 95% confidence interval, 4.71 to 24.87 percentage points). This interval pertains to the final cross-sectional contrast, not to a controlled estimate of the intervention's causal effect.

DISCUSSION

The results are consistent with the educational rationale for differentiated practice and formative feedback: students' programming performance involves multiple connected processes, and targeted assistance can be adjusted as those processes unfold. The AmaliyLab model extends simple task delivery by coupling difficulty adjustment with modulation of support and by keeping teachers responsible for

pedagogical interpretation and intervention.

The initial equivalence of aggregated categorical distributions strengthens the descriptive comparison; however, non-significant baseline chi-square results do not establish complete equivalence across all learning-related variables. The significant final association and the observed percentage-point improvement suggest potential educational value but do not, by themselves, demonstrate that AmaliyLab was the sole cause of the improvement. Instructor effects, selection processes, differences in engagement and implementation intensity remain plausible influences.

This revision deliberately separates two kinds of evidence. The original article described a Bayesian Knowledge Tracing simulation with 5,000 synthetic learners, but the dissertation's verified experimental dataset and pedagogical model do not establish that these simulation parameters or outputs characterize

the implemented AmaliyLab system. The simulation claims are excluded from the present empirical results to avoid conflating a hypothetical adaptive engine with the actual multi-institutional evaluation. If verifiable simulation code and an explicit technical mapping to AmaliyLab become available, a separate computational evaluation may be justified.

Additional limitations concern the use of grouped categorical data; unavailable learner-level paired scores; undocumented competency cutoffs, rater agreement and scale reliability; incomplete evidence on academic-year cohort overlap; and limited information on software implementation and recorded usage. Future work should report task-bank specifications, versioned software functionality, matched baseline/final student records, transparent allocation procedures and reproducible code for analysis.

CONCLUSIONS

AmaliyLab provides a structured adaptive pedagogical framework for developing students' practical programming activity through diagnosis, differentiated assignments, monitored practice, staged feedback, reflection and teacher-supported adjustment. In the dissertation-based, three-university dataset ($N = 307$), combined high/medium competency shares increased from 41.18% to 77.78% in the experimental groups and from 40.91% to 62.99% in the control groups. The 2×3 final distribution difference was statistically significant ($\chi^2 = 23.4514$; $df = 2$; $p \approx 0.0000081$; Cramér's $V \approx 0.2764$). The results support further examination of AmaliyLab-supported instruction, while a causal effectiveness claim requires complete individual-level data, technical documentation and stronger control of confounding factors.

DECLARATIONS AND RESEARCH TRANSPARENCY

Ethical approval, participant consent procedure, funding and conflicts of interest: information should be confirmed by the authors and participating institutions before journal submission. Data availability: the analyses in this article are based on aggregated tables in the dissertation; any release of anonymized learner-level data requires documentation and appropriate authorization. No unsupported software repository, simulation code or external dataset is claimed to accompany this revision.

REFERENCES

1. Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens. Publications Office of the European Union.
2. Decree of the President of the Republic of Uzbekistan No. PF-5847 (8 October 2019), "On approval of the Concept for the development of the higher education system of the Republic of Uzbekistan until 2030."
3. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
4. Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78.
5. Bloom, B.S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.
6. Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. Springer.
7. Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278.
8. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.