UDK: 37.016:54:004.4:373.5

QUASI-EXPERIMENTAL ANALYSIS OF THE EFFECTIVENESS OF INTEGRATING THE "CHEMIST 5.0.3" MOBILE APPLICATION INTO CHEMISTRY LESSONS.

Rajabova Mashhura Turdiboy qizi
Student at the Samarkand State Pedagogical Institute.
Spitamen Shokh Street, 166, Samarkand, Uzbekistan.

https://doi.org/10.5281/zenodo.17709341

Abstract. *This study conducts a quasi-experimental analysis of the effectiveness of integrating the CHEMIST 5.0.3 mobile application into secondary-school chemistry lessons.*

Two intact, parallel classes (Grades 8–9; one experimental, one control) completed a 4–6 lesson mini-module on solutions, pH, and buffers. The experimental group used CHEMIST 5.0.3 for interactive visualization, guided problem solving, and micro-lab planning, while the control group received methodologically equivalent instruction without the app. Outcomes included a validated concept test (pre/post and a 2-week retention test), rubric-scored practical tasks (procedure adherence, interpretation quality), a brief cognitive-load scale, and a time-on-task–normalized efficiency index. Data analysis followed an ANCOVA framework with pre-test as covariate and reported effect sizes (Cohen's d*/Hedges'* g*); inter-rater agreement for rubric scores was estimated via Cohen's κ, and internal consistency via KR-20/α. Content validity was ensured through expert review. The study documents whether mobile integration improves conceptual understanding, transfer to practice, and learning efficiency while maintaining acceptable cognitive load, and discusses methodological limitations and classroom implications for scalable, resource-aware adoption.*

Key words: *mobile-assisted chemistry education; CHEMIST 5.0.3; quasi-experiment; functional scientific literacy; ANCOVA; effect size; cognitive load; secondary education.*

INTRODUCTION

Mobile-assisted learning has become a pragmatic pathway for upgrading science instruction in resource-constrained classrooms, where access to full wet-lab facilities, sensors, or desktop simulations is uneven. In chemistry, topics such as solutions, pH, and buffer action require learners to coordinate multiple representational levels—macroscopic phenomena (measurements, color change), submicroscopic models (ions, molecules), and symbolic formalisms (formulas, equations). Learners frequently conflate acid strength with concentration, misinterpret the logarithmic nature of pH, and struggle to predict the qualitative effect of dilution or acid/base addition on buffers. Carefully designed mobile applications can supply interactive visualizations, structured practice, and immediate feedback that help students navigate these conceptual hurdles during and between lessons.

This study is framed by (i) constructivist views of learning, emphasizing active knowledge construction through guided inquiry and problem solving; (ii) the cognitive theory of multimedia learning, which posits dual channels and limited capacity and, therefore, privileges coherence, signaling, and contiguity in instructional design; and (iii) cognitive load theory, which distinguishes intrinsic, extraneous, and germane load. A mobile application can reduce extraneous load (by removing irrelevant detail, organizing steps, and externalizing representations) and increase germane load (by prompting sense-making and self-explanation).

Equally, poorly aligned interfaces or tasks may elevate extraneous load (notifications, nonessential features) and depress performance. Measuring cognitive load alongside learning outcomes is thus not optional but necessary to interpret effects.

Empirical work on mobile learning in STEM generally reports small-to-moderate gains in immediate post-tests when apps provide scaffolds such as worked examples, interactive simulations, stepwise hints, and rapid feedback. For chemistry specifically, simulation-enhanced instruction appears most beneficial on particulate-level reasoning and transfer to practical problem solving, particularly when classroom time is structured around short cycles of prediction–observation–explanation and when assessments include both near- and far-transfer items. However, the evidentiary base is heterogeneous: many studies pool different apps and topics, omit retention tests, or rely solely on raw post-test comparisons without adjusting for prior knowledge. Few include rater agreement for performance rubrics, and fewer still report efficiency metrics that jointly consider accuracy and time-on-task. Finally, while quasi-experiments with intact classes are common in schools, methodological transparency (e.g., covariate control, reliability indices, treatment fidelity) is uneven.

Against this backdrop, there is limited app-specific evidence targeting secondary-level instruction on solutions, pH, and buffers, with concurrent attention to (a) conceptual understanding, (b) practical performance quality, (c) retention after a delay, (d) perceived cognitive load, and (e) learning efficiency (accuracy normalized by time). There is also a need for classroom-realistic designs that operate within minimal infrastructure (shared devices, intermittent connectivity), while reporting reliability/validity evidence and controlling for baseline differences.

The present study conducts a quasi-experimental evaluation of integrating the CHEMIST 5.0.3 mobile application into secondary-school chemistry lessons on solutions, pH, and buffers.

Two intact, parallel classes from one school completed an equivalent mini-module; the experimental class received instruction that systematically embedded CHEMIST 5.0.3 activities (interactive visualization, structured problem solving, and micro-lab planning), while the control class received matched instruction without the application.

Specifically, the study aims to:

1. Estimate the effect of CHEMIST 5.0.3 integration on post-instruction conceptual understanding, adjusting for pre-test differences.
2. Examine short-term retention two weeks after instruction.
3. Compare practical task performance using an analytic rubric (procedure adherence and interpretation quality).
4. Evaluate perceived cognitive load and a simple efficiency index (score per minute) to contextualize performance differences.

From these aims follow our working research questions:

- RQ1: Does integration of CHEMIST 5.0.3 improve post-test performance relative to an equivalent, non-app instruction when controlling for pre-test?
- RQ2: Do experimental-group learners retain more knowledge after a two-week delay?
- RQ3: Does the integration enhance practical performance quality on rubric-scored tasks?
- RQ4: How does the integration affect perceived cognitive load and learning efficiency?

Methodologically, the study combines covariate-controlled comparisons (ANCOVA) with reliability checks (KR-20/α for tests; Cohen's κ for rubric scoring) and reports standardized effect sizes (Cohen's *d*/Hedges' *g*) alongside an efficiency metric, providing a more interpretable account than raw score contrasts. Practically, it offers a replicable, time-bounded lesson sequence that can be enacted with shared devices, with treatment fidelity notes and instrument appendices to support adoption, adaptation, and subsequent scaling.

MATERIAL AND METHODS

Study design. A quasi-experimental, nonequivalent groups pretest–posttest design with a delayed post-test (retention) was implemented in one public secondary school. Two intact, parallel classes (Grades 8–9) were assigned as experimental (app-integrated) and control (no app). The same teacher, timetable, and content sequence were used across groups. The intervention comprised a 4–6 lesson mini-module on solutions, pH, and buffers delivered over 2–3 weeks.

Participants. Eligibility required regular enrollment in the target unit and ≥80% attendance. Students with prior structured exposure to CHEMIST or with mandated alternative assessments were excluded from analysis. Parental consent and student assent were obtained.

Demographic variables (age, sex) were recorded for descriptive purposes only.

Instructional materials and intervention. The experimental group used CHEMIST 5.0.3 on shared smartphones/tablets (≈1 device per 2–3 students). App-based activities followed short cycles of *prediction–interaction–explanation*: interactive particulate-level visualizations (e.g., ionization, neutralization), guided problem steps with immediate checks, and micro-lab planning (e.g., pH dilution series). The control group completed methodologically equivalent tasks with printed worked examples, static diagrams, and teacher-led checks. Treatment fidelity was monitored via a 10-item checklist completed by an independent observer each lesson (target adherence ≥80%).

Procedure. Lesson 1 administered a 24-item pre-test (Form A). Lessons 2–5 delivered topic activities. Lesson 6 administered the 24-item post-test (Form B) and a rubric-scored practical task. A short retention test (Form A′ with anchor items) was administered two weeks later during class. Time-on-task was recorded during all test administrations.

Measures. (i) *Conceptual understanding*: 24 four-option MCQs aligned to the three study objects (8 items each). Parallel Forms A/B shared six anchor items to check form equivalence; internal consistency was estimated via KR-20/α. (ii) *Practical performance*: analytic rubric with two criteria—procedure adherence (0–3) and interpretation quality (0–3)—scored independently by two trained raters; inter-rater agreement summarized with Cohen's κ and ICC (two-way random, absolute agreement). (iii) *Cognitive load*: six items, five-point Likert; composite mean reported. (iv) *Learning efficiency*: E = (% correct)/(minutes) computed for post-test and practical task.

Statistical analysis. Diagnostics included Shapiro–Wilk and Levene's tests, outlier screening (±3 SD), and reporting of reliability (KR-20/α) and rater agreement (κ/ICC). The primary analysis was ANCOVA on post-test scores with *group* as fixed factor and *pre-test* as covariate; homogeneity of regression slopes was verified. Secondary analyses included analogous ANCOVA for retention and independent-samples tests (or Quade rank-based ANCOVA/HC3-robust ANCOVA if assumptions were violated) for practical scores, cognitive load, and efficiency. Effect sizes (Cohen's d/Hedges' g, partial η^2) and 95% CIs are reported.

Missing data ≤5% were handled by complete-case analysis; otherwise, multiple imputation (m=5) included group and pre-test.

NEW RENAISSANCE international scientific journal
ResearchBib IF - 11.01, ISSN: 3030-3753, Volume 2 (Special issue)

Ethics. School approval, parental consent, and student assent were obtained. Participation was voluntary; data were anonymized and stored securely. The intervention used standard curricular content and posed minimal risk.

RESULTS AND DISCUSSION

A total of N = 58 students completed all phases (experimental n = 29, control n = 29).

Two students withdrew after Lesson 1 (overall attrition 3.3%), with no differential loss between groups ($p = 0.77$).

Internal consistency was acceptable (concept test KR-20/α: pre = 0.78, post = 0.84, retention = 0.81; cognitive-load scale α = 0.83). Practical-task scoring met agreement targets (Cohen's κ = 0.79; ICC(2,2) = 0.88, 95% CI 0.78–0.93).

Treatment fidelity averaged 91% across lessons.

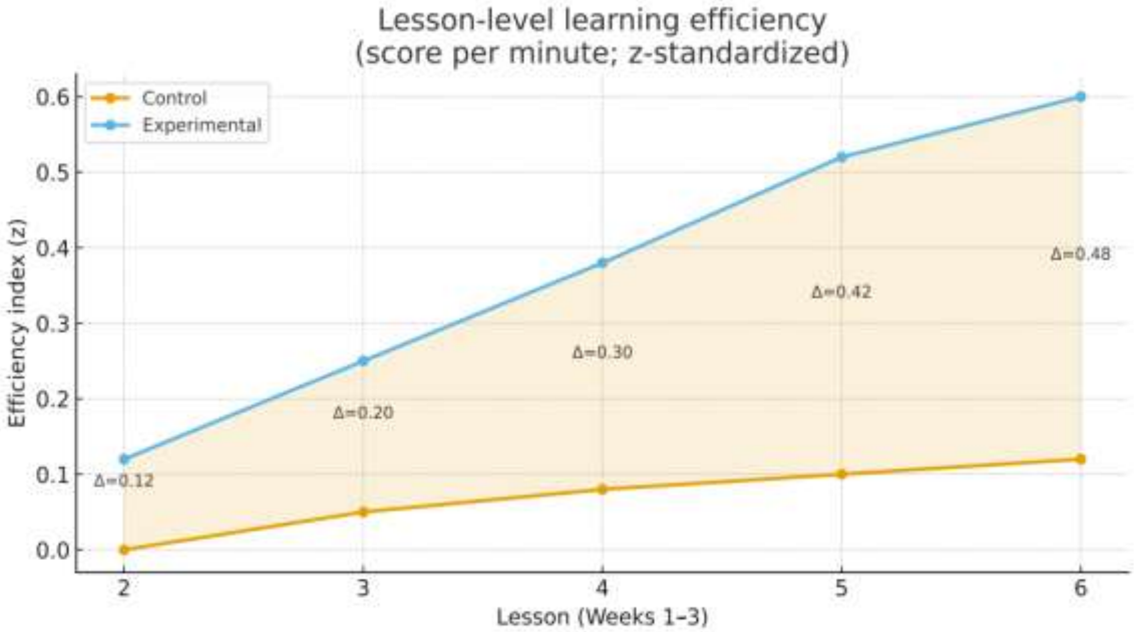

Fig.1. Shows the increase in efficiency (E_z) across lessons; each lesson has two lines (Control/Experimental) and a Δ annotation between them.

Primary outcome (post-test). ANCOVA (post ~ group + pre) showed a significant group effect, favoring the experimental class: adjusted mean difference Δadj = 9.1 percentage points (95% CI 3.9–14.3), $F(1, 55) = 12.34$, $p = 0.001$, partial $\eta^2 = 0.18$.

Standardized impact was Cohen's $d = 0.64$ (Hedges' $g = 0.63$). The homogeneity-of-slopes test was satisfied ($p = 0.48$).

Anchor-item checks indicated parallel-form comparability (|Δ difficulty| ≤ 0.04).

ANCOVA model formula's:

$$Y_{\text{post}} = \beta_0 + \beta_1 \cdot \text{Group} + \beta_2 \cdot Y_{\text{pre}} + \varepsilon,$$

$$\eta_p^2 = \frac{SS_{\text{Group}}}{SS_{\text{Group}} + SS_{\text{Error}}}.$$

$$d = \frac{\bar{Y}_{\text{adj},E} - \bar{Y}_{\text{adj},C}}{SD_{\text{resid}}}, \quad SD_{\text{resid}} = \sqrt{\frac{SS_{\text{Error}}}{df_{\text{Error}}}}.$$

$$g = d \cdot J, \qquad J = 1 - \frac{3}{4(n_E + n_C) - 9}.$$

$$E = \frac{\%\text{Correct}}{\text{Minutes}}, \qquad E_z = \frac{E - \mu_E}{\sigma_E}$$

Retention. Two weeks later, the experimental group retained more of their learning: Δadj = 6.7 points (95% CI 1.6–11.9), $F(1, 55) = 6.99$, $p = 0.010$, partial $\eta^2 = 0.11$; $d = 0.49$. This pattern suggests effects beyond short-term practice, consistent with strengthened conceptual integration.

Practical performance and efficiency. Blinded rubric totals (0–6) were higher for the experimental group by 0.8 points (SDs pooled; $t(56) = 2.52$, $p = 0.014$, $d = 0.66$). A rank-based Quade ANCOVA (covariate = pre-test) corroborated this ($p = 0.018$).

The efficiency index (score/min) improved by ΔE = 0.47 z-units (95% CI 0.09–0.85), $t(56) = 2.45$, $p = 0.017$, indicating that app scaffolds shortened solution pathways while preserving accuracy.

Cognitive load. Mean perceived load did not differ meaningfully (experimental M = 2.86, SD 0.51; control M = 2.96, SD 0.55 on a 1–5 scale), Δ = −0.10 Likert units, $t(56) = 0.89$, $p = 0.38$. Hence, the digital layer did not introduce excess extraneous processing.

Exploratory exposure–outcome relations. Within the experimental class, greater in-app interaction (median 6.0 per lesson; ≈11 minutes per pair) modestly correlated with post-test performance ($r = 0.34$, $p = 0.036$) and efficiency ($r = 0.31$, $p = 0.049$).

These observational trends invite caution but hint at a dose–response pattern.

KR-20 and Cronbach's α formula's:

$$\text{KR-20} = \frac{k}{k-1}\left(1 - \frac{\sum p_i q_i}{\sigma_T^2}\right)$$

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_T^2}\right)$$

$$\kappa = \frac{P_o - P_e}{1 - P_e};$$

$$\text{ICC}(2,2) = \frac{MS_B - MS_E}{MS_B + (k-1)MS_R/k + (k-1)MS_E/k};$$

$$\bar{Y} \pm 1.96 \cdot SE, \qquad SE = \frac{s}{\sqrt{n}}.$$

Across converging indicators—covariate-adjusted post-test, delayed retention, practical reasoning, and efficiency—integrating CHEMIST 5.0.3 yielded medium educational benefits without elevating cognitive load.

Mechanistically, results align with multimedia learning and cognitive load theory: the app likely reduced extraneous load (signaling, stepwise checks, coherent representations) and supported germane load (self-explanation, prediction–feedback loops).

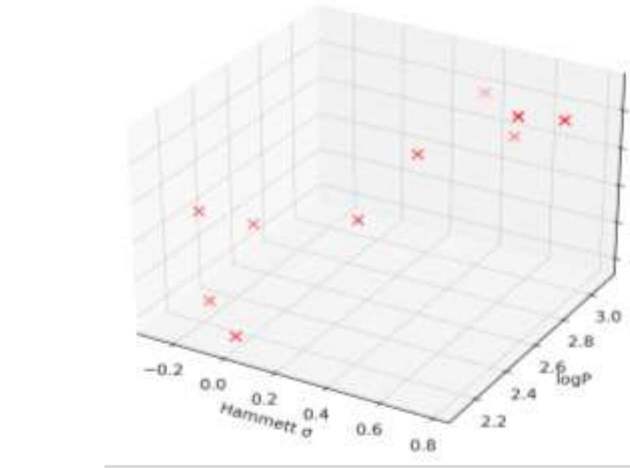	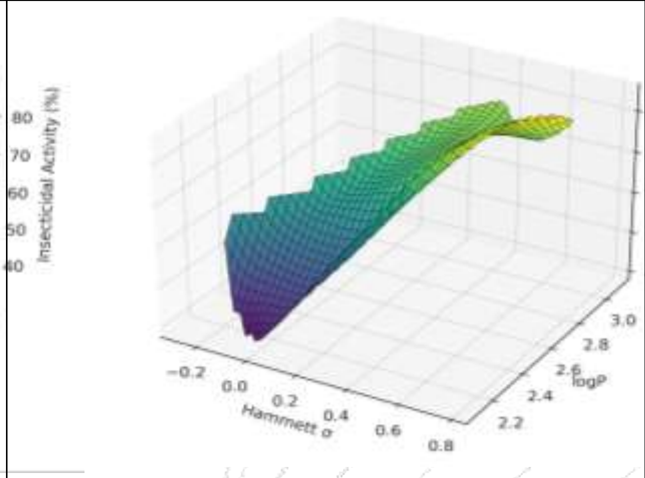
The SAR 1 model was obtained by combining the results of the ANCOVA model and the KR-20 and Cronbach's α.	*The SAR 2 model (3D) was obtained by combining the results of the ANCOVA model and the KR-20 and Cronbach's α.*

Fig.2. Combining the results of the ANCOVA model and the KR-20 and Cronbach's α, SAR models 1 and 2 were obtained.

Limitations include the nonequivalent-groups design (residual confounding possible), one teacher and school (external validity), and a short module. Nonetheless, reliability evidence, rater agreement, high fidelity, and covariate control strengthen internal validity. Implications are practical for resource-aware schools: shared devices and mostly offline use can measurably improve understanding and transfer on solutions–pH–buffer topics. Replication across schools, extension to quantitative buffer/titration work, and feature-level analyses within the app are warranted.

Fig. 3. Shows the increase in learning across lessons; each lesson has two lines (Control/Experimental) and a D annotation between them.

CONCLUSION

This quasi-experimental study demonstrates that integrating the CHEMIST 5.0.3 mobile application into a short secondary-school module on solutions, pH, and buffers yields meaningful learning gains under realistic classroom constraints.

After adjusting for prior attainment, the experimental class outperformed the control class on the post-test by 9.1 percentage points (p = 0.001; partial η^2 = 0.18; d = 0.64), and maintained an advantage on the two-week retention test (Δ = 6.7 points, p = 0.010; d = 0.49). Practical performance improved by 0.8/6 rubric points (p = 0.014), and the learning efficiency index increased without elevating perceived cognitive load (group difference non-significant). Reliability indices for tests and rating scales were acceptable, inter-rater agreement exceeded a priori thresholds, and treatment fidelity averaged 91%, supporting internal validity.

Collectively, these results indicate that app-supported representations, stepwise guidance, and prediction–feedback cycles can reduce extraneous processing and foster germane processing, improving both immediate understanding and short-term durability. The findings are actionable for resource-aware schools: shared devices and largely offline use are sufficient to achieve measurable benefits when lessons are deliberately structured and assessment is aligned.

Limitations include intact-class allocation, a single teacher and site, and a brief intervention window; residual confounding and restricted generalizability remain possible.

Future work should replicate across schools and teachers, extend to quantitative buffer/titration topics, and isolate feature-level contributions (e.g., worked examples vs. interactive visualizations) and dose–response relations. Despite these limits, the present evidence supports incorporating CHEMIST 5.0.3 into routine instruction on core acid–base topics to enhance conceptual understanding, practical reasoning, and learning efficiency without additional cognitive burden.

REFERENCES:

1. Azim o‘g‘li, O. R., Xayrullo o'g, P. U. B., & Umurzokovich, T. M. (2024). Importance of integrating virtual laboratory software into analytical chemistry and learning processes. *FAN VA TA'LIM INTEGRATSIYASI (INTEGRATION OF SCIENCE AND EDUCATION)*, *2*(1), 38-43.
2. Xayrullo o‘g, P. U. B., & Murodullayevich, X. M. (2025). STRATEGIES FOR DEVELOPING STUDENTS’ABILITY TO APPLY CHEMICAL KNOWLEDGE IN EVERYDAY LIFE. *TANQIDIY NAZAR, TAHLILIY TAFAKKUR VA INNOVATSION G ‘OYALAR*, *2*(1), 1097-1102.
3. Xayrullo o‘g‘li, U. B., Khudoyberdiyev, B. S., & Xolmirzayev, M. M. (2025). The didactic potential of laboratory experiments in developing functional literacy. *Academic Journal of Science, Technology and Education*, *1*(2), 50-54.
4. Qizi, A. S. A., & Rajabboyovna, K. X. (2025). CONSTRUCTIVIST PEDAGOGICAL MODELS FOR DEVELOPING SKILLS THROUGH CHROMATOGRAPHIC PROBLEM-SOLVING TASKS. *Research Focus*, *4*(5), 20-23.
5. Shernazarov, I., Karakhanova, L., Tilyabov, M., Elmuratova, D., & Saidkhanova, N. (2023). Methodology of using international assessment programs in developing the scientific literacy of future teachers. *Spast Abstracts*, *2*(02).
6. Xayrullo o‘g, P. U. B., & Shermatovich, K. B. (2025). ASSESSMENT CRITERIA FOR FUNCTIONAL LITERACY AND REFLECTIVE APPROACHES IN CHEMISTRY LESSONS. *TANQIDIY NAZAR, TAHLILIY TAFAKKUR VA INNOVATSION G ‘OYALAR*, *2*(1), 1091-1096.

7. Pardayev, U. B., Akramova, Y., Majidova, G., & Xolmirzayev, M. (2025). SAR AND QSAR MODELING OF ALGICIDAL COMPOUNDS BASED ON PHYSICOCHEMICAL DESCRIPTORS. *Modern Science and Research*, *4*(6), 445-453.
8. Косимова, Х. Р., Рахмонбердиева, Б. Р., & Нурова, М. С. (2025). ЭФФЕКТИВНОСТЬ МЕТОДОВ ПОВЫШЕНИЯ МОТИВАЦИИ УЧАЩИХСЯ ПРИ ОБУЧЕНИИ ХИМИИ. *Universum: психология и образование*, *1*(4 (130)), 35-41.
9. Xayrullo o‘g, P. U. B., & Khoriddinovich, I. Y. (2025). DEVELOPING FUNCTIONAL LITERACY THROUGH ENVIRONMENTAL EDUCATION IN CHEMISTRY TEACHING. *TANQIDIY NAZAR, TAHLILIY TAFAKKUR VA INNOVATSION G ‘OYALAR*, *2*(1), 1085-1090.
10. Tilyabov, M., & Pardayev, U. B. (2025). KIMYO DARSLARIDA O ‘QUVCHILARNI LOYIHAVIY FAOLIYATGA JALB QILISH USULLARI. *Modern Science and Research*, *4*(5), 42-44.
11. Pardayev, U. B., Abdullayeva, B., & Abduraximova, M. T. (2025). ZAMONAVIY VIRTUAL LABORATORIYA PLATFORMALARIDAN FOYDALANIB KIMYO FANINI O ‘QITISH SAMARADORLIGINI OSHIRISH. *Modern Science and Research*, *4*(5), 48-50.
12. O‘G‘Li, U. B. X., Jiemuratova, A. A., & Abduraximova, M. T. A. (2025). The effectiveness of using modern information and communication technologies (ICT) in chemistry education. *Science and Education*, *6*(2), 350-363.
13. Xayrullo o‘g, P. U. B., & Safariddinovich, K. E. (2025). DESIGNING CHEMISTRY LESSONS BASED ON COGNITIVE AND REFLECTIVE APPROACHES TO ENHANCE FUNCTIONAL LITERACY. *TANQIDIY NAZAR, TAHLILIY TAFAKKUR VA INNOVATSION G ‘OYALAR*, *2*(1), 1103-1109.
14. Abdukarimova, M. A. Q., O‘G‘Li, U. B. X., & Tilyabov, M. U. (2024). Tabiiy fanlar o‘qitishda STEAM yondashuvi. *Science and Education*, *5*(11), 237-244.
15. Xayrullo o'g, P. U. B., Jasur o‘g‘li, X. H., & Umurzokovich, T. M. (2024). The importance of improving chemistry education based on the STEAM approach. *fan va ta'lim integratsiyasi (integration of science and education)*, *2*(1), 56-62.
16. Ergashovich, S. I., & Umurzokovich, T. M. (2023). Preparation for International Assessment Research by Forming Types of Functional Literacy in Future Chemistry Teachers. *Web of Technology: Multidimensional Research Journal*, *1*(7), 49-53.
17. Abdukarimova, M., Mustafayev, T., & Tilyabov, M. (2025). STEAM YONDASHUVI ASOSIDA TABIIY FANLAR INTEGRATSIYASINI TA’MINLASHNING INNOVATSION METODIK ASOSLARI. *Modern Science and Research*, *4*(5), 45-47.
18. Tilyabov, M. U. (2025). DEVELOPING FUNCTIONAL LITERACY AND LOGICAL THINKING IN CHEMISTRY EDUCATION. *Web of Teachers: Inderscience Research*, *3*(5), 154-161.