BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

MEASURING THE EFFECT OF CREDIT-MODULE-BASED QUALITY MANAGEMENT: A THREE-DIMENSIONAL «3Q» APPROACH TO EVALUATING EDUCATIONAL CHANGE

Nigora N. Kurbonova
Independent Researcher, Tashkent State University of Economics, Tashkent, Uzbekistan
e-mail: niga.turinova@mail.ru
https://orcid.org/0009-0009-8200-2256
doi:https://doi.org/10.5281/zenodo.23036617

Abstract. Studies of quality management in higher education commonly report effect size along a single dimension — the average level of achievement reached. This paper argues that a single-dimension report conceals two managerially decisive properties of an educational change: how many students it reaches and how fast the change accumulates. A three-dimensional evaluation framework, termed here the «3Q» approach (Quality, Quantity, Quickness), is proposed and tested. A three-year quasi-experimental study (2021–2024) was conducted at three higher education institutions in Uzbekistan with 375 undergraduate students (195 experimental, 180 control) in the Economics field of study. Readiness for education quality management was measured across four criteria — motivational, emotional, informational and process-activity — each operationalised through six indicators on a three-point scale and converted into a level index ranging from 1 to 3. The experimental factor was a bundle of five pedagogical conditions built on module passports, criterion-referenced rubrics, hybrid-form selection driven by a motivational-emotional parameter, and threshold-triggered monitoring. Baseline equivalence was confirmed (between-group index differences ≤ 0.006). Post-test differences were statistically significant for every criterion at every institution (Student's t, $p < 0.05$). The three «3Q» components diverged sharply: Quality differed by a factor of 1.34, Quantity by 6.1 and Quickness by 6.0. The divergence is the central finding — reporting Quality alone would have described the intervention as moderate, whereas the coverage and rate dimensions show its actual scale. Growth was consistent across the three institutions (0.560–0.619), which supports transferability. Limitations and directions for further research are discussed.

Keywords: education quality management; credit-module system; quasi-experiment; level index; effect measurement; internal quality assurance; higher education in Uzbekistan; learning analytics.

1. Introduction

The transition of higher education systems to credit-based organisation of the learning process has changed the unit of educational management. Across the European Higher Education Area this transition is now largely complete, and attention has shifted from the introduction of credit instruments to the question of how their effects are to be verified [17]. Where the academic subject and the contact hour once served as the objects of planning and control, the module and the credit unit now occupy that position. This shift has been widely documented, yet its managerial consequence is less often examined: as the object of management becomes smaller and its cycle shorter, the instruments used to evaluate managerial interventions must become correspondingly more precise.

The difficulty begins with the construct itself. Quality in higher education is conventionally defined as the correspondence between stated purpose and attained result [18], which means that any claim about a change in quality is in fact a claim about a change in that correspondence — and therefore requires that both terms be measured. A recurring feature of the empirical literature on quality management interventions, however, is that their effect is reported along a single dimension — most commonly the mean level of achievement attained by the experimental group relative to a control group. Such a report answers one question: how high did the participants get. It does not answer two further questions that determine whether an intervention is worth institutionalising: how many participants were actually moved, and how quickly the movement accumulated. Two interventions may produce an identical mean and yet differ fundamentally — one may have lifted a small group of already-strong students, the other may have moved the bulk of a weak cohort.

This distinction is not merely descriptive. In a credit-module environment the module is short, and a managerial correction must be applied within its duration or not at all. An intervention whose effect accumulates slowly is, in operational terms, unusable at module level regardless of the mean it eventually produces. Rate of change therefore has independent managerial value and cannot be inferred from the final level.

The context of the present study adds a normative dimension. In Uzbekistan, Resolution No. 498 of the Cabinet of Ministers of 6 August 2025 introduced a system of comprehensive and specialised state accreditation in which internal (self-)evaluation was defined for the first time as a mandatory stage, described in the accompanying Regulation as a process of observing, analysing, evaluating and taking corrective measures with respect to an institution's own performance [1]. The National Agency for Quality Assurance in Education, established by Presidential Resolution No. PQ-275 of 8 September 2025, was charged, among other functions, with developing recommendations on internal quality control criteria [2]. Internal measurement thus acquired formal status, while the question of what precisely should be measured, and how the measured effect should be reported, was left to institutions.

The purpose of this paper is to propose and empirically test a three-dimensional framework for evaluating the effect of a quality-management intervention in a credit-module environment. The research questions are: (RQ1) does a bundle of credit-module-based pedagogical conditions produce a statistically significant change in students' readiness for education quality management; (RQ2) is that change consistent across institutions with different regional and disciplinary profiles; and (RQ3) do the three proposed dimensions — level attained, coverage and rate — behave differently, and if so, what does the divergence imply for how such effects should be reported.

2. Literature review

Research on internal quality assurance has increasingly turned from the design of systems to the question of whether they operate in practice. Aldhobaib, in an ethnographic study based on 53 interviews, found that formally established quality assurance systems frequently remain at the level of documentation and do not reach the teaching process itself [3]. The finding is significant for effect measurement: if a quality system can exist without producing change, then the presence of the system cannot be treated as evidence of its effect, and independent measurement becomes necessary.

Greere and Riley, working with eleven European institutions, propose a five-stage operational model of quality culture development and identify the transparency of assessment criteria — their being known to students in advance — as the stage at which perceptible change begins [4]. A systematic review by Berkat covering 22 studies published between 2010 and 2025 similarly concludes that internal quality assurance produces observable effects mainly where stakeholders

are engaged in the evaluation process rather than being its objects [5]. The European standards for quality assurance point in the same direction, orienting external review towards an institution's capacity to measure and respond to its own results rather than towards the results themselves [10]. Vargas and colleagues, reporting on the standardisation of competency-based course assessment across twenty courses and 109 instructors, show that such capacity depends on assessment criteria being defined at the level of the individual learning outcome [16].

A second strand concerns the credit unit itself. The ECTS methodology defines the credit in terms of the workload a student must actually invest rather than in terms of teaching hours delivered [11]. Andrade raises the question of whether declared credit volume corresponds to students' actual study time and argues that the correspondence should be verified rather than assumed [6]. Rivadeneyra provides empirical evidence linking ECTS credit allocation, real workload and teaching quality [7]. Both lines of work imply that a credit-module intervention must be evaluated against student effort rather than against nominal hours.

A third strand addresses the informational basis of quality management. Javed and Alenezi describe a data-warehouse and KPI-dashboard architecture for institutional quality monitoring [8]. Hershkovitz, Ambrose and Soffer, surveying 253 instructors, identify the binding constraint not in the collection of indicators but in the absence of a mechanism converting them into managerial decisions [9]. The observation is directly relevant here: an evaluation framework is useful only to the extent that its dimensions correspond to decisions an institution can actually take.

Regarding the reporting of effect, the dominant convention in the field remains a comparison of group means, supplemented in stronger designs by a significance test. Multidimensional reporting is uncommon. Where coverage is reported, it usually appears as a descriptive remark rather than as a distinct evaluation dimension with its own indicator. The framework proposed below is intended to fill that gap by treating level, coverage and rate as three co-equal dimensions of a single result.

3. Methodology

3.1. Design and participants

A quasi-experimental pre-test/post-test design with non-equivalent control groups was employed over three academic years (2021–2022 preparatory; 2022–2023 and 2023–2024 formative and summative). Three higher education institutions participated: Tashkent State University of Economics (TSUE), Jizzakh Polytechnic Institute (JizPI) and Fergana Polytechnic Institute (FerPI). The institutions were selected purposively: all three deliver the same field of study (Economics, by sectors and branches) and operate under the same credit-module regulation, while differing in regional location, institutional profile and group size. Table 1 reports the composition of the sample.

Table 1

Composition of the sample

Institution	Profile	Total	Experimental	Control
Tashkent State University of Economics	Economics	150	80	70
Jizzakh Polytechnic Institute	Technical	115	65	50
Fergana Polytechnic Institute	Technical	110	50	60

Institution	Profile	Total	Experimental	Control
Total	—	375	195	180

Source: compiled by the author.

3.2. Instruments and measurement

Readiness for education quality management was measured across four criteria. Two are parametric — motivational and emotional — and are measured before the learning process in order to serve as input parameters for the selection of the instructional form. Two are outcome criteria — informational and process-activity — and measure the effect of the managerial intervention. The choice of criteria follows established theory: the motivational criterion draws on self-determination theory [12], the emotional criterion on control-value theory [13], the informational criterion on the taxonomy of cognitive outcomes [14], and the process-activity criterion on the learning-analytics premise that participation is best measured from the records activity itself leaves rather than from self-report [15].

Each criterion was operationalised through six indicators scored on a three-point scale (1 = low, 2 = medium, 3 = high), giving a range of 6 to 18 points per criterion and placing each student at one of three levels: high (14–18), medium (10–13) or low (6–9). The motivational and emotional criteria were measured by closed-item questionnaires with pre-scored response options; the informational criterion by a two-part test separating recall of concepts from their application in a professional context; and the process-activity criterion from documented records — submission dates, participation in interim control points, and the distribution of activity in the electronic environment — recorded independently by two observers. Instruments were piloted with students not participating in the experiment.

To express the distribution of levels as a single quantity, a level index was computed as the weighted mean of level frequencies, with weights of 3, 2 and 1 for the high, medium and low levels respectively. The index ranges from 1 to 3, where 1 corresponds to all students at the low level and 3 to all at the high level. The same instruments were used at pre-test and post-test without modification.

3.3. The experimental factor

The experimental factor was not a single method but a bundle of five pedagogical conditions applied in the experimental groups while the control groups followed the usual arrangement of the learning process. The conditions were: (i) grouping of subjects in a logical sequence and purposeful modularisation, implemented through a module passport that records the learning outcome, credit volume and assessment form in a single document; (ii) selection of the hybrid instructional form on the basis of a motivational-emotional parameter measured at the start of each module; (iii) functional linkage of subject content and credit units on the principle of planning — control — transformation; (iv) use of an open criterion-referenced assessment rubric presented to students before the module begins; and (v) continuous monitoring of module outcomes with predefined thresholds triggering revision of module content or credit volume.

A verification indicator was assigned to each condition in advance of the intervention. This design choice constrains selective reporting: it fixes, before data collection, which indicator will be taken as evidence for which condition.

3.4. Data analysis and validity

Level distributions were computed as percentages; distributions were reduced to the level index; and between-group differences at post-test were tested using Student's t-test at the 0.05

significance level. Ninety-five per cent confidence intervals were computed for the post-test index in each group. Baseline equivalence was verified by comparing pre-test indices.

Typical threats to internal validity were addressed in the design. Baseline non-equivalence was tested statistically. Natural maturation was accounted for by recording growth in the control group separately. Instrumentation change was excluded by using identical instruments at both measurement points. Observation effects were mitigated by announcing in advance that questionnaire results would not affect grades and by deriving the process-activity criterion from documented records. Attrition was handled by including in the final analysis only students present at both measurements. Assessor subjectivity was limited by the use of closed items with pre-scored options and by dual independent recording of activity.

4. Results

4.1. Baseline equivalence

At pre-test both groups were dominated by the low level, which accounted for 68.3–73.3 per cent of students depending on the criterion. Between-group differences in the level index did not exceed 0.006 on a scale with a range of two units, that is, less than 0.3 per cent of the range. The groups may therefore be treated as equivalent at the outset, and post-test differences may be attributed to the experimental factor rather than to initial disparity.

4.2. Change in the distribution of levels

By post-test the experimental group had shifted markedly. The share of students at the low level fell from 68.7–73.3 per cent to 20.5–21.5 per cent, while the share at the medium level more than tripled. In the control group the decline of the low level did not exceed nine percentage points, and the low level continued to predominate. Figure 1 presents the shift for both groups.

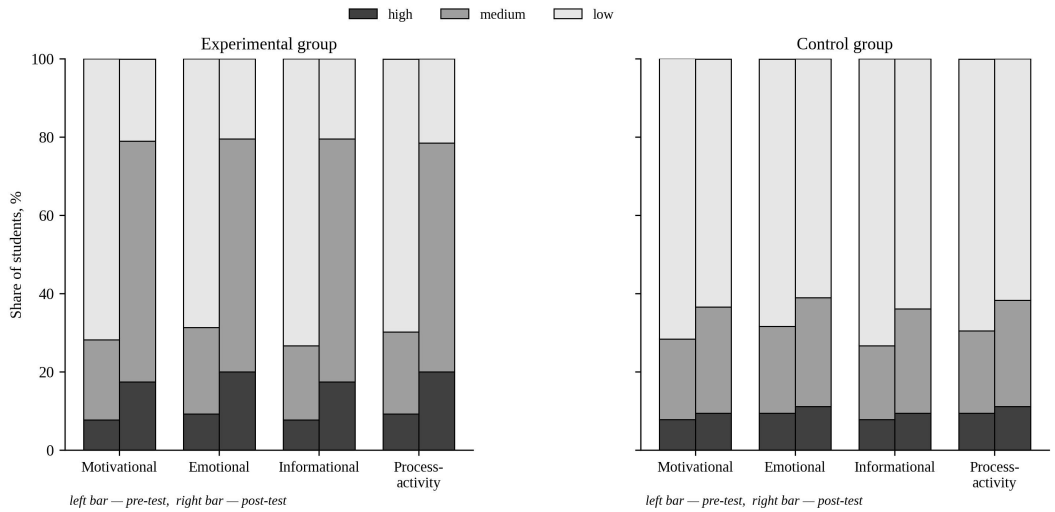


Figure 1. Distribution of levels at pre-test and post-test, by group and criterion

Source: computed by the author.

The direction of movement is informative. The bulk of students leaving the low level moved to the medium level rather than to the high level. This is consistent with the duration of the intervention: two academic years are sufficient to move a student out of the low level but not, for most, to reach the highest level of readiness.

4.3. Level index and statistical significance

Reduced to the level index, the experimental group rose from 1.375 to 1.975, a growth of 0.600, equal to 30 per cent of the index range. The control group rose from 1.378 to 1.478, a

growth of 0.100 — six times smaller. Growth was nearly uniform across the four criteria (0.58–0.63), indicating that the experimental factor did not act selectively on any single criterion. Table 2 reports the index dynamics together with the results of significance testing.

Table 2

Level index dynamics and Student's t-test results (aggregate, n = 375)

Criterion	Exp. pre	Exp. post	Growth	Control pre	Control post	t (df = 373)
Motivational	1.36	1.96	0.60	1.36	1.46	7.59
Emotional	1.41	1.99	0.58	1.41	1.50	7.22
Informational	1.34	1.97	0.63	1.34	1.46	7.77
Process-activity	1.39	1.98	0.59	1.40	1.49	7.11
Mean	1.375	1.975	0.600	1.378	1.478	—

Source: computed by the author. Critical value of t at df = 373 and p = 0.05 is 1.97; all empirical values exceed it.

Significance was confirmed not only in the aggregate but at each institution taken separately, where degrees of freedom were 148 (TSUE), 113 (JizPI) and 108 (FerPI) and empirical t values ranged from 3.65 to 5.09 against a critical value of 1.98.

4.4. Consistency across institutions

An aggregate result may conceal an effect confined to a single site. Index growth was therefore computed for each institution separately (Figure 2). Growth in the experimental groups ranged from 0.560 to 0.619 — a spread of 0.059, less than three per cent of the index range. The effect of the experimental factor is therefore not appreciably dependent on the regional location or the institutional profile of the site, which strengthens the case for transferability.

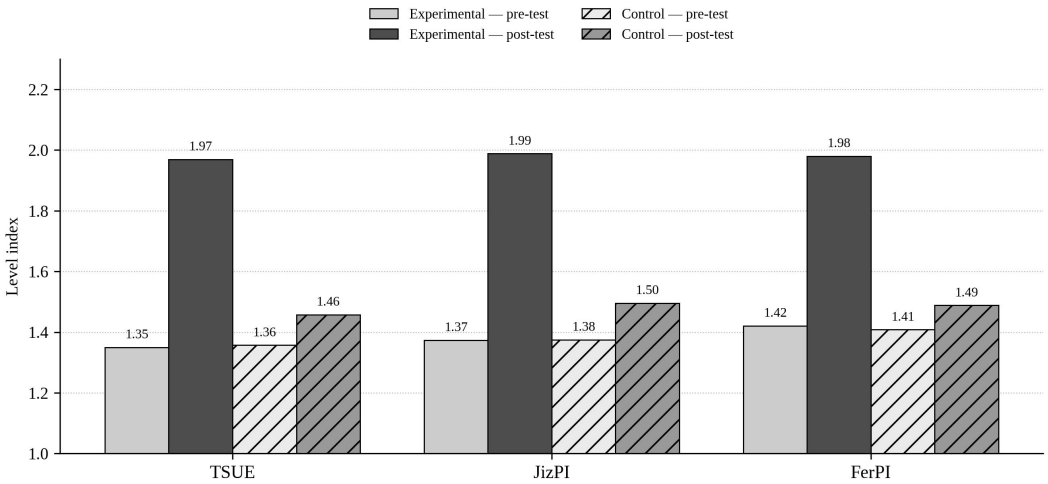


Figure 2. Level index by institution, pre-test and post-test

Source: computed by the author.

One pattern within the institutional data deserves comment. Fergana Polytechnic Institute began with the highest baseline index (1.420) and recorded the smallest growth (0.560). The inverse relation is expected: the higher the starting point, the smaller the headroom remaining

within a bounded scale. That post-test indices at the three sites converged within a narrow band (1.969–1.988) is explained by the same mechanism.

4.5. Confidence intervals

Significance testing establishes that a difference is unlikely to be due to chance but says little about its magnitude. Ninety-five per cent confidence intervals were therefore computed for the post-test index (Table 3). For every criterion the intervals of the two groups are disjoint, with a substantial gap between them; the narrowest separation occurs on the emotional criterion, where 0.304 index units — more than fifteen per cent of the scale range — lie between the upper bound of the control group and the lower bound of the experimental group.

Table 3

Post-test level index: means, standard deviations and 95 % confidence intervals

Criterion	Group	n	Mean	SD	SE	95 % CI
Motivational	Experimental	195	1.964	0.621	0.044	1.877 – 2.051
	Control	180	1.461	0.663	0.049	1.364 – 1.558
Emotional	Experimental	195	1.995	0.638	0.046	1.905 – 2.084
	Control	180	1.500	0.689	0.051	1.399 – 1.601
Informational	Experimental	195	1.969	0.617	0.044	1.883 – 2.056
	Control	180	1.456	0.663	0.049	1.359 – 1.552
Process-activity	Experimental	195	1.985	0.646	0.046	1.894 – 2.075
	Control	180	1.494	0.689	0.051	1.394 – 1.595

Source: computed by the author.

Standard deviations are similar in the two groups (0.617–0.689). The experimental factor therefore did not markedly increase within-group dispersion, which indicates that the result was produced by a general shift of the cohort rather than by gains concentrated among a minority of students. Had growth been confined to a subgroup, dispersion in the experimental group would have widened appreciably.

4.6. The three «3Q» dimensions

The analyses above establish that a difference exists and that it is robust, but they describe it along one dimension. The «3Q» framework decomposes the result into three: Quality, taken as the final level index; Quantity, taken as the share of students who left the low level, expressed relative to the group; and Quickness, taken as index growth per academic year. Figure 3 reports the three components.

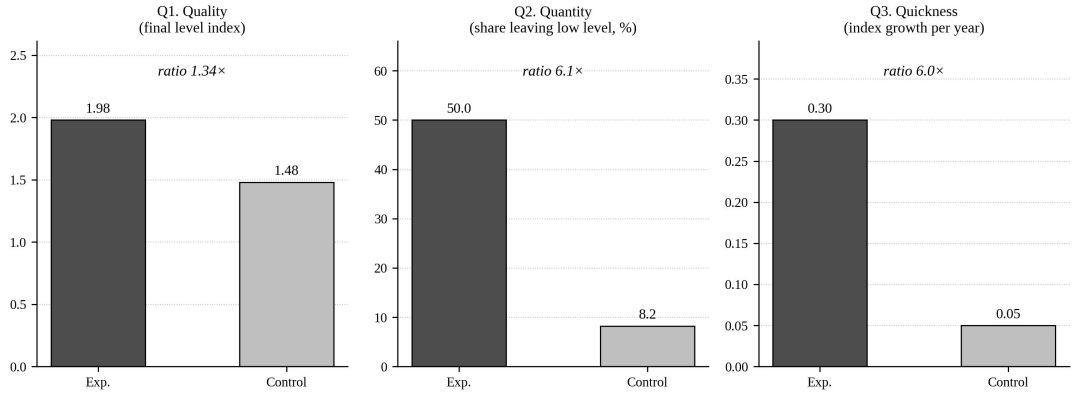


Figure 3. The three «3Q» components: experimental and control groups

Source: computed by the author.

The three components diverge sharply. Quality differs by a factor of 1.34, Quantity by 6.1 and Quickness by 6.0. The divergence is not an artefact but a consequence of the properties of the measures. The level index is bounded above at 3 and the control group also grew, so a ratio computed on it is compressed towards unity. Coverage and rate are not bounded in the same way and consequently register the difference at close to its actual magnitude.

The implication for reporting is direct. Had this study reported Quality alone, the intervention would have been described as producing a moderate effect of roughly one-third. The coverage dimension shows that half of the experimental cohort left the low level while fewer than one in twelve did so in the control group; the rate dimension shows that the experimental group accumulated in one academic year what the control group accumulated in six. Both facts are managerially decisive and both are invisible in the single-dimension report.

5. Discussion

The results speak to each of the three research questions. On RQ1, the intervention produced a statistically significant change on every criterion at every site, with non-overlapping confidence intervals. On RQ2, growth was consistent across three institutions differing in region and profile, which supports transferability within the same field of study. On RQ3, the three dimensions behaved differently, and the difference was large enough to change the substantive interpretation of the result.

The finding that the control group also improved, though six times less, is consistent with Aldhobaib's observation that formally functioning quality systems produce limited practical effect [3]. Both control and experimental sites operated under the same national quality-assurance framework; what distinguished them was not the presence of a quality system but whether that system reached the credit and module level of the learning process. The difference in outcome is therefore attributable to the depth at which quality management was implemented rather than to its existence.

That the highest post-test index was recorded on the emotional criterion (1.995) is consistent with Greere and Riley's identification of criterion transparency as the point at which quality culture becomes perceptible [4]. The open assessment rubric, presented before each module, was one of the five conditions tested, and the emotional criterion was its designated verification indicator.

The process-activity result (index 1.985, growth 0.59) bears on the workload question raised by Andrade [6] and Rivadeneyra [7]. Because this criterion was derived from documented records rather than self-report, its movement in the same direction as the self-reported criteria provides a

partial internal check: had the questionnaire-based criteria risen while activity records remained flat, the result would have been open to a response-bias interpretation.

The threshold mechanism used in the fifth condition addresses the constraint identified by Hershkovitz and colleagues, namely the absence of a route from indicator to decision [9]. Predefining the value at which a module's content or credit volume is revised converts monitoring from observation into action, and it is this element rather than the collection of data that distinguishes the experimental arrangement.

One result diverges from the international literature. Effects of credit-system implementation are commonly assessed as moderate, whereas the coverage dimension here differs by more than a factor of six. Two circumstances account for this. The baseline was unusually low, with 68–73 per cent of students at the lowest level, leaving substantial headroom; and the experimental factor was a bundle of five conditions rather than a single instrument. Both circumstances counsel caution in transferring the magnitude of the effect to settings with higher baselines or narrower interventions.

6. Conclusion

This study proposed and tested a three-dimensional framework for evaluating the effect of quality-management interventions in a credit-module environment. Four conclusions follow.

First, a bundle of five credit-module-based pedagogical conditions produced a significant change in students' readiness for education quality management: the level index rose from 1.375 to 1.975 against 1.378 to 1.478 in the control group, with all differences significant at $p < 0.05$ and confidence intervals disjoint on every criterion.

Second, the effect was consistent across three institutions with different regional and disciplinary profiles, with growth ranging from 0.560 to 0.619 — a spread below three per cent of the scale range.

Third, the three evaluation dimensions diverged substantially: 1.34 for Quality, 6.1 for Quantity and 6.0 for Quickness. Reporting the first alone would have understated the intervention; the practical implication is that studies of this kind should report coverage and rate alongside attained level.

Fourth, similar within-group dispersion in both groups indicates that the change reflects a general shift of the cohort rather than gains concentrated among stronger students.

Three limitations qualify these findings. The experimental factor was applied as a bundle, so the contribution of its individual components cannot be separated; isolating them requires a factorial design. The study was conducted within a single field of study, and transfer to other fields requires verification, since the nature of learning outcomes and consequently the parameters of hybrid forms differ by discipline. Finally, feedback quickness was measured indirectly, as the rate of index growth, rather than directly as elapsed time from submission to feedback; direct measurement requires a dedicated field in the institutional information system, which was not available during the period of the study.

Further research will accordingly proceed in three directions: separating the contribution of individual pedagogical conditions; testing the framework in other fields of study; and implementing direct measurement of feedback time, which would allow the Quickness dimension to be reported on its own scale rather than as a derived rate.

References

1. Aldhobaib M.A. Quality assurance struggle in higher education institutions: moving towards an effective quality assurance management system // Higher Education. – 2024. – Vol. 88, No. 4. – P. 1547–1566.
2. Javed Y., Alenezi M. A Case Study on Sustainable Quality Assurance in Higher Education // Sustainability. – 2023. – Vol. 15, No. 10. – Article 8136.
3. Kurbonova, N. P. (2022). PRINCIPLES OF AN E-COURSE DESIGN. Архив научных исследований, 2(1).
4. Hershkovitz A., Ambrose G.A., Soffer T. Instructors' Perceptions of the Use of Learning Analytics for Data-Driven Decision Making // Education Sciences. – 2024. – Vol. 14, No. 11. – Article 1180.
5. Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). – Brussels: European Students' Union, 2015. – 32 p.
6. ECTS Users' Guide 2015. – Luxembourg: Publications Office of the European Union, 2015. – 108 p.
7. Kurbonova, G. (2022). TA'LIM JARAYONIDA O 'QITISHNING MODUL TIZIMI TAMOYILLARI. Архив научных исследований, 2(1).
8. Sultanova, D. T., & Muratova, M. N. (2022). THE ROLE OF COMPUTER TECHNOLOGY AS A RESOURCE FOR FORMATION LEARNING ENVIRONMENT FOR STUDYING AND TEACHING ENGLISH LANGUAGE. Архив научных исследований, 2(1).
9. Deci E.L., Ryan R.M. Intrinsic Motivation and Self-Determination in Human Behavior. – New York: Plenum Press, 1985. – 371 p.
10. Pekrun R. The Control-Value Theory of Achievement Emotions // Educational Psychology Review. – 2006. – Vol. 18, No. 4. – P. 315–341.
11. Siemens G., Long P. Penetrating the Fog: Analytics in Learning and Education // EDUCAUSE Review. – 2011. – Vol. 46, No. 5. – P. 30–40.
12. Nematovna, Q. N. (2023). Improving the management of the quality of education by introducing the credit module system in higher education institutions. Eurasian Journal of Learning and Academic Teaching, 20, 8891.
13. Vargas H., Arredondo E., Heradio R., de la Torre L. Standardizing course assessment in competency-based higher education: an experience report // Frontiers in Education. – 2025. – Vol. 10. – Article 1579124.
14. European Commission / EACEA / Eurydice. The European Higher Education Area in 2024: Bologna Process Implementation Report. – Luxembourg: Publications Office of the European Union, 2024.
15. Harvey L., Green D. Defining Quality // Assessment and Evaluation in Higher Education. – 1993. – Vol. 18, No. 1. – P. 9–34.