Journal of Vasyl Stefanyk Precarpathian National University
http://journals.pnu.edu.ua
Vol. 13, No. 1 (2026), 145-160

doi: 10.15330/jpnu.13.1.145-160

ISSN 2311-0155 (Print)
ISSN 2413-2349 (Online)

ARCHITECTURE OF ASSESSMENT SYSTEMS IN HIGHER EDUCATION: MULTILEVEL MODEL AND DIAGNOSTIC CRITERIA

ANTON ZAPOROZHETS, TETIANA ZAPOROZHETS*

*Corresponding author: ant.zaporozhets@gmail.com

Abstract. Assessment systems in higher education are typically described through scales, points, and formal procedures. However, can one understand what function assessment actually performs – supporting learning or administrative control – by relying solely on the type of scale or tools? Why do formally similar systems in different universities lead to fundamentally different educational practices and student behavioral strategies? Assessment cannot be adequately analyzed through individual elements, since identical solutions can perform different functions depending on the mechanism of their formation, combination, and use. The purpose of the article is to develop an architectural analytical framework for designing and diagnosing assessment systems in higher education as a multilevel configuration of interrelated decisions. To achieve this, heterogeneous characteristics of assessment systems have been systematized as decision variants at individual architectural levels, typical system configurations for different institutional purposes have been modeled, and diagnostic criteria for analyzing the constructive coherence of a specific assessment system have been developed. The scientific novelty lies in the proposed architectural approach: the conceptual level defines the space of variants for the methodological level, which is modified by regulatory parameters and implemented through organizational forms. Typical assessment trajectories and diagnostic tools for evaluating their coherence are presented. The model operationalizes the principle of balance between pedagogically desirable, technically feasible, and administratively permissible, to design systems for mindful learning even under institutional constraints. The main conclusion: assessment systems are valid only under the condition of internal architectural coherence between levels.

Keywords*:* higher education, constructive alignment, formative assessment, assessment coherence, assessment design.

1. INTRODUCTION

Assessment became a systemic tool only in the 18th–19th centuries with the beginning of mass education and the demand for standardization of educational outcomes. Before this, the process was regulated informally: through individual dialogue and oral disputations, where the result was recognition of qualification, not a fixed grade. In Europe, the first steps toward quantitative formalization were recorded in 1792 at Cambridge University, when student performance was first mathematically measured and administratively compared using a system of numerical indicators (Hoskin, 1979, p. 144). In the American tradition, assessment was initially based on a ranking approach, according to which a grade depended not on achieving an absolute criterion of knowledge, but on the student's position in the group. The transition to descriptive forms occurred at Yale University (1785), and later at Harvard University (1877) division into statistical "divisions" was introduced (Durm, 1993, p. 295). Standardized letter notations (A–F) were finally established only at the end of the 19th century

(particularly at Mount Holyoke College in 1897), when the need arose to select large numbers of students (Durm, 1993, p. 297).

Thus, assessment methodology from the very beginning developed in managerial interests as an instrument of ranking and control. However, before the era of mass education, particularly in ancient and medieval traditions, assessment was carried out through individual dialogue and was oriented toward recognition of acquired mastery, not comparison between students. The current shift in educational paradigm is an attempt to return to this primary orientation toward individual development: to shift focus from comparing students with each other to pedagogical support of each student. Such a shift involves creating conditions for mindful learning, when a student understands their goals, receives feedback about progress, and can adjust their trajectory, and the assessment system becomes an instrument of this support, not just control. This is confirmed by the concept of formative assessment (Sadler, 1989, p. 120; Andriichuk, 2023, p. 152), the movement for alternative assessment systems, for example specifications grading (Nilson & Packowski, 2025, pp. 6, 140), as well as the tendency to reduce the role of high-stress forms of control (Howitz et al., 2025, p. 18). The result is unstable tension between declared pedagogical goals and actual assessment mechanisms that retain their standardized, administratively-oriented nature. John Biggs drew attention to the tension between administrative and academic requirements back in 1996 and called on academics to "become more proactive, positively insisting that educational considerations should prevail over administrative convenience" (Biggs, 1996b, p. 361).

A review of scientific literature sources on assessment in higher education indicates that existing research focuses either on individual types of scales and grade notations, or on specific instruments and practices (peer assessment (Yin et al., 2022)), or on psychometric aspects (grade inflation (Lipnevich et al., 2020, p. 16)), which are typically analyzed outside institutional or systemic context. For example, a large-scale review (Brookhart et al., 2016), covering more than 100 years of assessment research, captures the multidimensionality of scales and variability of practices, but does not propose a generalized framework that could answer the question: why does one and the same form of assessment work differently in different contexts, and how do systemic parameters influence the pedagogical or administrative character of assessment.

Recent systematic reviews reveal gaps between the declared principles of innovative assessment practices (authentic assessment, self/peer assessment, ungraded learning) and their actual implementation (Mesny et al., 2025, p. 3). Research on faculty competencies identifies the multilevel nature of assessment systems – macro- (program), meso- (course), and micro- (assignment) levels, each requiring specific decisions (Boud & Dawson, 2023, pp. 162–163). Empirical studies demonstrate contradictions between declared formative goals and regulatory parameters in examinations (Remesal & Estrada, 2023, p. 12). They also reveal that formal system design does not fully determine student practices, as students make evaluative judgments through informal interactions (Fischer et al., 2024, p. 241).

Thus, despite recognition of the multilevel nature of assessment and identification of gaps between goals and practices, heterogeneous characteristics of systems (scales, mechanisms, regulatory parameters) are studied fragmentarily. This highlights the need for their conceptual analysis and systematization to clarify the nature of elements and principles of their coherent integration.

2. THEORETICAL FRAMEWORK

Analysis of international experience and domestic assessment practices reveals that assessment systems cannot be holistically described exclusively through typology of scales or forms of assessment. A multidimensional system with different natures of categories, different levels of decisions, different assessment goals is traced. A series of questions arises that takes assessment components beyond the boundaries of linear classification of concepts.

Is ranking-based assessment a separate system, or is it just an algorithm for determining a grade? It turns out that ranking-based assessment is one of ***the grade formation mechanisms*** through comparison of results between students, which depends on the distribution of results in the group, for example, grading on a curve, when the top 10 % receive the highest grade, even if their absolute score is not high (Mehvar, 2025). There are two other grade formation mechanisms. The criterion-referenced mechanism is based on comparing a student's learning achievements with predefined criteria, for example, a student receives 85 points if they demonstrated 85 % of planned competencies. The normative-threshold mechanism determines the final grade not only by learning outcome, but also by compliance with procedural requirements not directly related to the quality of learning outcome (for example, mandatory number of points for current work to be admitted to the exam, regardless of performance quality).

The next question concerns how to represent the result. Most often in the sections of normative documents of higher education institutions, "scales" of assessment and rules for converting grades between them are mentioned, for example, "100-point scale", "ECTS scale". However, it turns out that these are not different scales, but different notations of the same ordinal scale. It should be noted that modern ECTS since 2009 does not define a fixed grading scale (European Commission, 2015, p. 39), which creates terminological confusion in the practices of Ukrainian HEIs – a detailed analysis of the genesis of ECTS and quasi-ECTS configurations will be presented in a separate work. ***A grading scale*** is an ordered set of values for representing learning achievements. All scale forms in higher education are ordinal in the sense of educational measurement, that is, they show only order (who is better, who is worse), while intervals between grades are not proportional to intervals between actual learning outcomes. Dichotomous (Pass/Fail), verbal ("excellent / good / satisfactory"), letter (A–F) and numerical (0–100, 1–10) forms are notations of an ordinal scale, not separate types of scales. Since ordinal scales do not guarantee equal intervals between values, performing summation or averaging of grades is incorrect. For example, 75 points is considered as the average between 70 and 80, although in reality the difference in knowledge may be uneven (Sadler, 2009, p. 171). Therefore, when performing arithmetic operations on grades, it is assumed that numerical notations are used in a quasi-interval mode for aggregation (averaging points, calculating GPA), which does not change their ordinal nature, but allows management decisions to be made (Brookhart et al., 2016, p. 20). Narrative forms of assessment do not use a formalized scale at all – instead they provide individualized comments about student progress without an ordered set of values. Hybrid systems combine scale (with different notations) and non-scale (narrative) elements. That is why at the level of the architectural model, the term ***result representation form*** is used, which covers both scale and non-scale methods of representing learning achievements. One and the same result representation form can be used in different grade formation mechanisms and be embedded in systems with different goals and functions. For example, numerical notation 0–100 can be applied in both criterion-referenced assessment (a student receives 85 points for achieving defined criteria) and ranking-based (85 points corresponds to the top 10 % of the group after result normalization).

Which systems are suitable for formative assessment, and which for summative? This question requires distinguishing ***the temporal logic*** of assessment. Cumulative logic involves gradual formation of a grade over time with the possibility of providing intermediate feedback and correction of learning actions. Point-in-time (snapshot) logic captures the result at a specific moment without possibility of changes. It is important that ranking-based systems are procedurally incompatible with cumulative logic, since a student's position in ranking can be determined only after completing assessment of all group participants.

If the range of grades, assessment criteria, algorithms for accumulating and averaging points are described in assessment regulations, then a whole series of constraints are not related to the assessment system at all and are not recorded in the institution's official documents. Conditions such as "admission to exam – 36 points", "on retake maximum 75 points", "for late assignment half the points" and similar ones are "hidden" in syllabi of individual disciplines, methodological recommendations, or remain

verbal agreements of the academic. Are such constraints an element of the assessment system? Do they affect the objectivity of learning outcome assessment? Do they affect the learning process and student motivation? Yes, and significantly. Such additional rules are ***regulatory parameters*** of the assessment system: minimum thresholds for admission to control measures (for example, mandatory 36 points for current work), weights of individual assessment components (for example, 60 % current + 40 % exam), penalties (for example, reduction of maximum score on retake), restrictions (for example, prohibition on retaking if the grade is below threshold). They can both support the pedagogical function of the system: structure the learning process, ensure rhythmic work. And they can deform it: replace support of learning with sanctioning control, lead to non-objective assessment, provoke students to develop strategies for overcoming assessment barriers instead of focusing on learning. Thus, regulatory parameters perform the role of a "hidden algorithm" that can radically change the functions of the assessment system.

Overall, the assessment system encompasses not only methodological implementation, components of which are described above, but also requires alignment with goals and principles of assessment at one pole and with forms of organizational implementation and ways of using results at the other. ***Assessment goals*** determine why assessment is conducted: to support the learning process, control achievements, select students, certify qualifications, or for management reporting. ***Assessment principles*** determine the reference base for comparison: individual student progress (ipsative), achievement of predefined criteria (criterion-referenced), or position among other students (norm-referenced). Assessment results depend on ***the form of organizational implementation*** (assessment is conducted as formative at the level of individual assignments/modules or as summative after completion of course study, or as programmatic for determining qualification level) and the ***way results are used*** (internal institutional scales, inter-institutional instruments of academic recognition, national qualification frameworks and reporting instruments).

Is there a fundamental difference between pedagogical and administrative approaches to building an assessment system? Yes, and this distinction is the key to understanding assessment functions. Pedagogically-oriented systems are aimed at supporting learning (assessment for learning): they support accumulation of results, allow retaking without penalties, provide feedback for trajectory correction. Such systems create conditions for mindful learning, when a student can track their progress, understand success criteria, and make informed decisions about their own educational trajectory. Administratively-oriented systems are intended for management decisions (assessment for decision-making): they record the final result, are convenient for selection, ranking and reporting. It is worth noting that in each category described above that is part of the assessment system, one can distinguish decisions that tend toward pedagogical or administrative direction. For example, the narrative result representation form (gives detailed feedback) is "assessment for learning" (Wiliam, 2017) versus numerical notation, which is ideal for statistical processing in decision-making. However, numerical notation can be applied in both cumulative formative assessment (pedagogical function) and point-in-time ranking-based selection (administrative function). Thus, the essence of the assessment system is determined not by individual decisions of different natures, but by their coherent combination to achieve a goal defined by an individual academic or institutional educational policy.

Structuring decision variants from different categories as a list or table does not allow establishing connections between them. Therefore, a hierarchical structure is needed that reflects the sequence of interrelated decisions and their coherence between structural levels. The combination of elements into multilevel constructions suggests an analogy with architectural design. Each level of assessment architecture (assessment architecture, grading architecture) defines the space of permissible variants for the next, and a specific assessment system is formed through sequential selection and alignment of elements between levels, particularly between the extreme levels of declared goals and actual implementation. If declared goals contradict regulatory parameters or the way results are used, an *architectural gap* arises that deforms the system's function.

A similar principle of internal coherence (constructive alignment) was proposed by Biggs (1996b) for curriculum design and applied by Boud and Falchikov (2006) to assessment systems. The architectural model proposed in this work extends this logic, including regulatory parameters and administrative mechanisms as equal elements of assessment architecture.

In the second half of the 20th century, the term washback (backwash) emerged, which described the impact of testing and assessment on teaching and learning. The washback effect can be positive, stimulating learning, or negative, narrowing the content of the program, as academics focus only on test tasks. The washback of quantitative assessment methods pushes academics toward superficial approaches in teaching, which "typically lead to low cognitive-level outcomes that are not compatible with stated course objectives" (Biggs, 1996a, p. 5). In essence, the washback effect demonstrates that what is assessed is often perceived as valuable, and therefore – this is what is taught. Therefore, "assessment provides a strategic point of intervention to influence the development of learning for the longer term" (Boud & Falchikov, 2006, p. 405). However, John Biggs notes that not only goals influence decision-making, but also formal and informal requirements of the social system. Therefore, a "balancing the desirable, the possible and the allowable in tertiary education" is needed (Biggs, 1996a, p. 12): "what is likely" (existing assessment systems) is the result of compromise between "what we want" (goals) and "what is allowable" (what our administrations and colleagues will permit and implementers will be able to do).

The questions and conceptual distinctions outlined above form the basis for developing an architectural model of assessment systems.

3. RESEARCH OBJECTIVE, METHODOLOGY AND DATA

The purpose of this article is to develop an architectural analytical framework for designing and diagnosing assessment systems in higher education as a multilevel configuration of interrelated decisions. To achieve this goal, it is necessary to: systematize heterogeneous characteristics of assessment systems as decision variants at individual architectural levels; model typical system configurations for different institutional purposes; develop diagnostic criteria for analyzing the constructive coherence of a specific assessment system.

The research is based on an architectural approach that considers assessment as a multilevel configuration of decisions and involves analysis of constructive alignment between architectural levels. Methodologically, the research combines: (1) conceptual analysis to distinguish concepts that are often mixed in the literature (mechanism, scale, assessment system) and define architectural levels; (2) comparative analysis of assessment practices in higher education in different countries to identify typical system configurations; (3) modeling of design trajectories for different institutional goals; (4) development of diagnostic tools for quantitative-qualitative analysis of assessment system coherence.

4. RESULTS AND DISCUSSION

The multidimensionality of assessment systems examined in the theoretical framework (Section 2) – heterogeneous grade formation mechanisms, result representation forms, temporal logic, regulatory parameters – demonstrates that a system cannot be described through individual elements or linear typology.

Summarizing the above, the ***assessment system*** emerges as a multilevel architecture of interrelated decisions: the conceptual alignment (goals and principles) defines the space of variants for the methodological implementation (result representation forms, temporal logic, grade formation mechanisms), which is modified by regulatory parameters and implemented through organizational levels. Each architecture level defines the space of permissible variants for the next. A specific system is formed through sequential selection and alignment of elements. If declared goals contradict regulatory parameters or the way results are used, an architectural gap arises that deforms the pedagogical or

administrative function of the system.

The proposed architectural model synthesizes the ideas of constructive alignment (Biggs, 1996b), the washback effect, and measurement theory, operationalizing them as an analytical framework for diagnosing internal coherence of assessment systems. Unlike existing typologies that classify systems by individual features (type of scale, form of control), the proposed approach reveals hidden functions of assessment systems through analysis of coherence between architectural levels.

4.1. Architectural Model of Assessment Systems

The model reflects ***the architecture of assessment systems*** in higher education as a multilevel configuration of interrelated decisions (Fig. 1):

- conceptual alignment of goals (why assess?) and principles (what to compare and with what?);
- methodological implementation through result representation form (how to represent the result?), temporal logic (when and how to form a grade over time?) and grade formation mechanism (by what rule do we transform results into a grade?); representation form and temporal logic mutually determine each other (indicated by double arrows on the diagram); regulatory parameters can operate between temporal logic and grade formation mechanism, influencing permissible decisions at the next level;
- organizational implementation through assessment form and institutional results.

Each level defines the space of permissible variants, and a specific assessment system is formed through sequential selection and alignment of elements between levels.

Regulatory parameters are not a separate architectural level, but operate between levels as modifiers and can significantly change the assessment trajectory, similar to how an optical lens can change the path of a ray. In Fig. 1, the influence of regulatory parameters is indicated by lightning bolts – the number of bolts symbolizes the significance of regulatory influence. Critical is not the very presence of regulatory parameters, but their coherence with declared assessment goals. One lightning bolt on the blue trajectory means minimal regulation to ensure compatibility between systems, which does not create internal contradiction. Two lightning bolts on the pink trajectory correspond to strict thresholds and normalization, which are natural for the administrative purpose (selection), therefore also do not deform the trajectory. In contrast, three lightning bolts on the red trajectory signal an architectural gap: the same sanctioning mechanisms (thresholds, penalties, restrictions) come into conflict with the pedagogical purpose (supporting learning), transforming the system into a quasi-formative configuration, where formative assessment rhetoric masks actual administrative control.

Conditions for obtaining assessment results (temporal logic) must be compatible with their representation forms, so corresponding decisions are made in mutually coordinated fashion (indicated by double arrows on the diagram). For cumulative assessment, which involves sequential formation of a grade over time with the possibility of providing intermediate feedback, the most suitable are numerical notations and narrative forms, while letter notations require intermediate conversion to numerical equivalent. Point-in-time (snapshot) assessment is compatible with numerical and letter notations, and can be combined with any grade formation mechanism (criterion-referenced, normative-threshold, ranking-based), whereas narrative forms of assessment in this case are difficult to standardize. Thus, temporal logic does not determine result representation form directly, but narrows the space of permissible representational decisions and may require reconsideration of the initially chosen form of presenting assessment in the process of system design.

Assessment architecture has an internal pedagogical-administrative gradient: left decision variants at each level are oriented primarily toward supporting learning, while right ones – toward management tasks of control, selection and certification.

Within the proposed architecture, a fundamental distinction is made between the assessment system itself as a pedagogical mechanism for forming and interpreting learning achievements and the secondary management use of its results (discussed in more detail in paragraph 4.4).

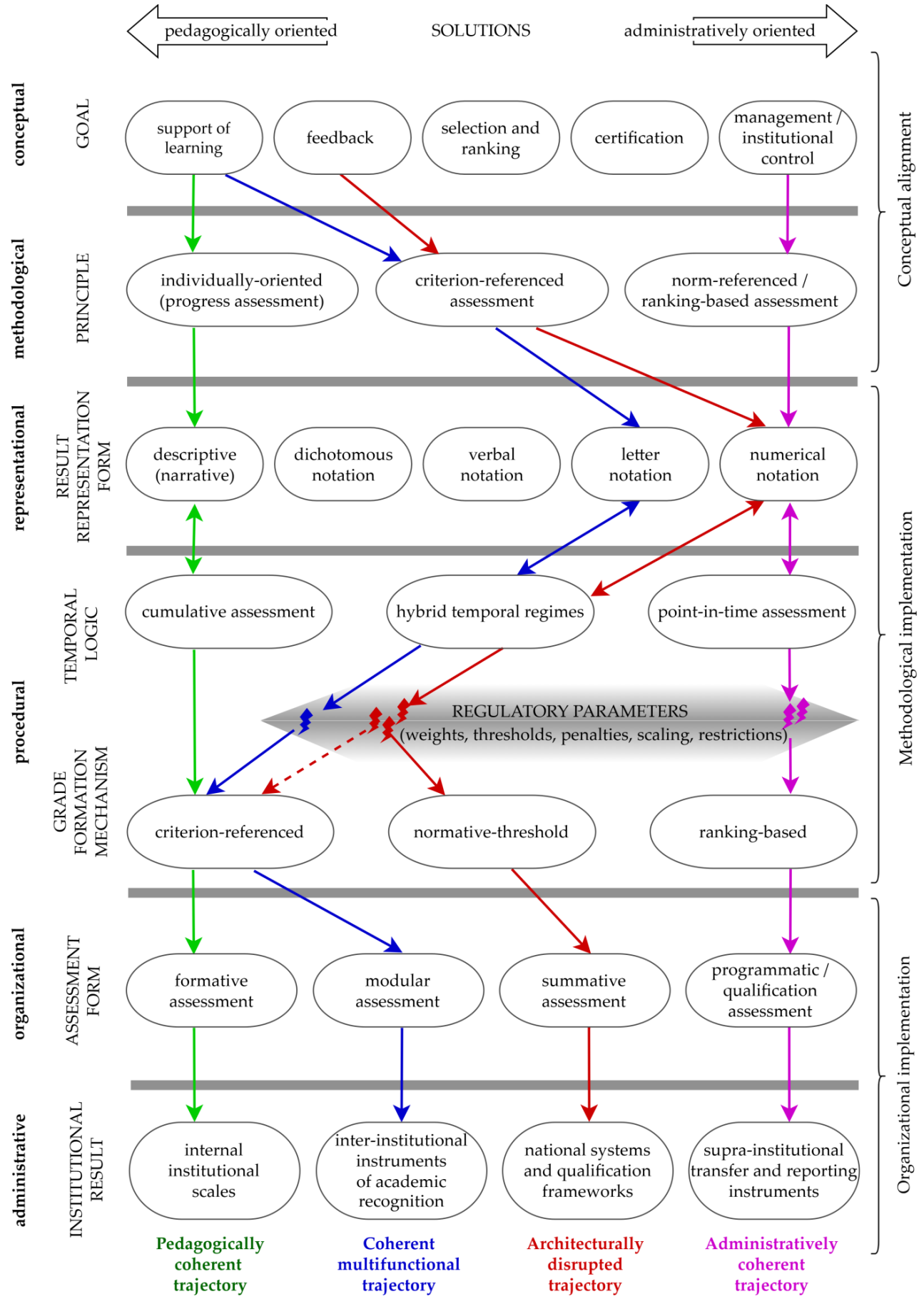

Fig. 1. Generalized Model for Designing Assessment Systems in Higher Education. Arrows indicate illustrative assessment trajectories

Source: Created by authors

4.2. Typical Design Trajectories

The proposed diagram allows designing various assessment models depending on set goals, as well as predicting behavioral and psychometric effects as a consequence of the chosen configuration. Below are illustrated ***four typical trajectories*** representing individual pedagogically and administratively-oriented scenarios and not claiming to exhaustively describe all possible configurations.

Formative Assessment with Portfolio (green trajectory). The trajectory on the left of the diagram represents a pedagogically-oriented assessment scenario aimed at maximum support of learning. In this configuration, assessment goals are focused on forming feedback and supporting individual progress, which is aligned with individually-oriented principles of progress assessment, narrative forms and cumulative temporal logic. The grade formation mechanism remains criterion-referenced without intervention of regulatory parameters. Organizationally, such a model is implemented as formative assessment, and the administrative level is limited to internal institutional scales without ranking and selection.

Key feature: the trajectory completely avoids regulatory parameters, preserving focus on individual student progress without comparison with other students or normative thresholds. Such a configuration maximally supports mindful learning: the student receives detailed feedback about their strengths and weaknesses, can review and improve work, independently determines the pace and priorities of learning. This is an internally consistent model in which all levels are subordinate to pedagogical logic. Examples of implementation of this trajectory: narrative grades in universities of applied sciences in Finland and Norway, portfolio approach in Liberal Arts colleges in the USA (Hampshire College, n.d.; Evergreen State College, n.d.), narrative transcripts (New College of Florida, n.d.), individualized learning trajectories (Minerva Project, n.d.).

Recognition of Prior Learning (blue trajectory). The trajectory illustrates a coherent multifunctional model where pedagogical and administrative goals are combined without internal contradictions. It describes a productive scenario for recognition of prior learning, microcredits, or certification of achievements. Assessment preserves criterion-referenced logic and orientation toward achieving learning outcomes through letter notations and hybrid temporal regimes, but experiences weak influence of regulatory parameters (one lightning bolt on the diagram), which ensure minimum thresholds for recognition and rules for conversion between systems. The grade formation mechanism remains criterion-referenced. Organizationally implemented through modular assessment, and results are used primarily for recognition and transfer through inter-institutional instruments of academic recognition.

Key feature: the administrative level plays a supporting role, ensuring legitimacy and comparability without interfering with the learning logic. Regulatory parameters here are natural and necessary to ensure compatibility between systems, but do not deform pedagogical goals. Examples: transfer of academic credits between universities within the Bologna space (KU Leuven), recognition of non-formal learning outcomes (RPL) in Canadian colleges (Fanshawe College, 2022), microcredits and short-term programs with accumulation in European universities (University College Dublin, n.d.), digital competency badges.

Cumulative System with Administrative Dominance (red trajectory). An internally contradictory trajectory, which in many university documents is declared as formative assessment, but is actually used for administrative control. It begins with pedagogically-oriented goals of supporting learning and criterion-referenced principles, uses numerical notations and hybrid temporal regimes, but at the procedural level is significantly modified through intensive influence of regulatory parameters (three lightning bolts on the diagram). Strict thresholds, penalties for retaking, maximum grade limitations, high weight of final control transform the grade formation mechanism from criterion-referenced to normative-threshold. Organizationally implemented as summative assessment, and administratively integrated into national systems and qualification frameworks.

Key feature: the dotted line on the diagram visualizes trajectory deformation through regulatory parameters. Such settings change student behavior, causing demotivation, increase in number of appeals, formalization of learning activity, and focus on maximizing points instead of achieving learning outcomes. Such a configuration blocks mindful learning: instead of focusing on their own development, the student is forced to develop strategies for overcoming assessment barriers, and the assessment system is perceived as an unpredictable source of threats, not an instrument of support. The gap between declared pedagogical goals and actual administrative implementation creates systemic contradiction that undermines trust in assessment. Examples: typical practice of Ukrainian universities, where summative control retains a dominant role not so much through formal weight (40–50 %), but through regulatory parameters: strict exam admission thresholds, penalties for retaking, maximum grade limitations on retakes.

Standardized Exams and Competitive Selection (pink trajectory). The trajectory on the right of the diagram corresponds to an administrative-management scenario. Assessment goals from the very beginning are related to control, selection and certification; the methodological basis is norm-referenced or ranking-based assessment; representation is through numerical notations; temporal logic is point-in-time. Regulatory parameters (two lightning bolts on the diagram) include passing scores and result normalization, which are natural and expected for this strategy and do not create internal contradiction. The grade formation mechanism is ranking-based. Organizationally implemented through programmatic or qualification assessment, and administratively – through supra-institutional transfer and reporting instruments.

Key feature: an internally coherent model in which all levels are subordinate to administrative logic. Regulatory parameters here do not deform the trajectory, since the assessment function is fundamentally different from the pedagogically-oriented one – the goal is not supporting learning, but ensuring selection and comparability of results on large samples. Examples: National multi-subject tests (ZNO, NMT) in Ukraine, competitive entrance examinations to selective universities, licensing exams for professional qualifications (medical, legal), international standardized tests (SAT, GRE, GMAT).

4.3. Diagnostic Tools

The diagram can be used as an analytical tool for designing and reviewing assessment policies: trajectories along the pedagogical (green) or administrative (pink) poles reflect internally coherent models, while diagonal trajectories can be productive hybrids (blue) or create systemic contradictions (red) and require particularly careful alignment between levels.

For quantitative diagnosis of assessment architecture, five criteria have been defined (Tab. 1), which form two parameters: strategic orientation of the system S (criteria 1–4) and internal coherence C (criterion 5). This allows analytically distinguishing viable configurations from institutionally unstable ones regardless of the terminology used.

Each of criteria 1–4 is evaluated as +1 (pedagogical feature) or –1 (administrative feature). If within one criterion different parameters are present or the system has mixed characteristics, it is necessary to determine the dominant tendency and choose the corresponding feature. For example, for criterion 2: if a cumulative system (60 %) is combined with summative control (40 %), but intermediate results really influence the final grade and correction is possible – this is a pedagogical feature (S + 1); if in fact everything is decided by the exam – administrative (S – 1). If there are several regulatory parameters (thresholds, penalties, restrictions) that help structure learning, then by criterion 3 we add (S + 1), if they predominantly limit and sanction, then we subtract (S – 1).

The strategy parameter S is calculated as the sum of evaluations by criteria 1–4, so it can take values from –4 to +4.

By criterion 5, the coherence parameter C evaluates correspondence between declared assessment goals and actual logic of its implementation through regulatory parameters and organizational form. External administrative use of results (scholarship rankings, competitive selection) is not an element of

the assessment system as such and does not create an architectural gap if it does not deform the internal mechanism of grade formation. An architectural gap arises when declared pedagogical goals contradict the sanctioning character of regulatory parameters or when administrative goals are implemented through non-transparent or inappropriate mechanisms.

Tab. 1

Five diagnostic criteria of two-parameter diagnosis (S, C) of assessment architecture

Strategy parameter S		
Criterion	Pedagogical feature: add one unit S + 1	Administrative feature: subtract one unit S – 1
1. Declared goal from educational program, syllabus (conceptual level)	supporting learning progress, providing feedback, mindful learning	selection and ranking, control and reporting, resource management
2. Temporal logic of assessment (procedural level)	result is formed gradually, possibility of correcting learning actions is provided	summative control actually determines the grade (through weight and/or regulatory constraints that nullify the influence of current work)
3. Regulatory parameters and grade formation mechanism (procedural level)	have auxiliary character, sanctioning constraints do not change the semantic logic of assessment	significantly correct or limit learning outcome, sanctions dominate, not feedback
4. Assessment form (organizational level)	at the level of individual assignments, modules as a tool of current feedback	at the level of program or qualification as a certification tool
Coherence parameter C		
Criterion	Coherent strategy: C = 1	Internally contradictory strategy: C = 0
5. Coherence (conceptual and administrative levels)	results are used according to declared function, no contradictions between levels	results are systematically used not for initial purpose, persistent tension between goals and practices

Source: Created by authors

Two-parameter diagnosis (S, C) allows analytically distinguishing viable configurations from institutionally unstable ones (Fig. 2). The boundaries –2 and +2 are defined as thresholds of clear predominance of one of the strategies. At these values, at least three of four criteria (i.e., absolute majority of architectural levels) have orientation of one type, which allows identifying the dominant system strategy: at $S \geq +2$, at least three of four criteria have pedagogical orientation, at $S \leq -2$ – administrative. The range $-2 < S < +2$ corresponds to systems with balanced or differently oriented

features, which can be both productive multifunctional configurations (at C = 1) and internally contradictory (at C = 0).

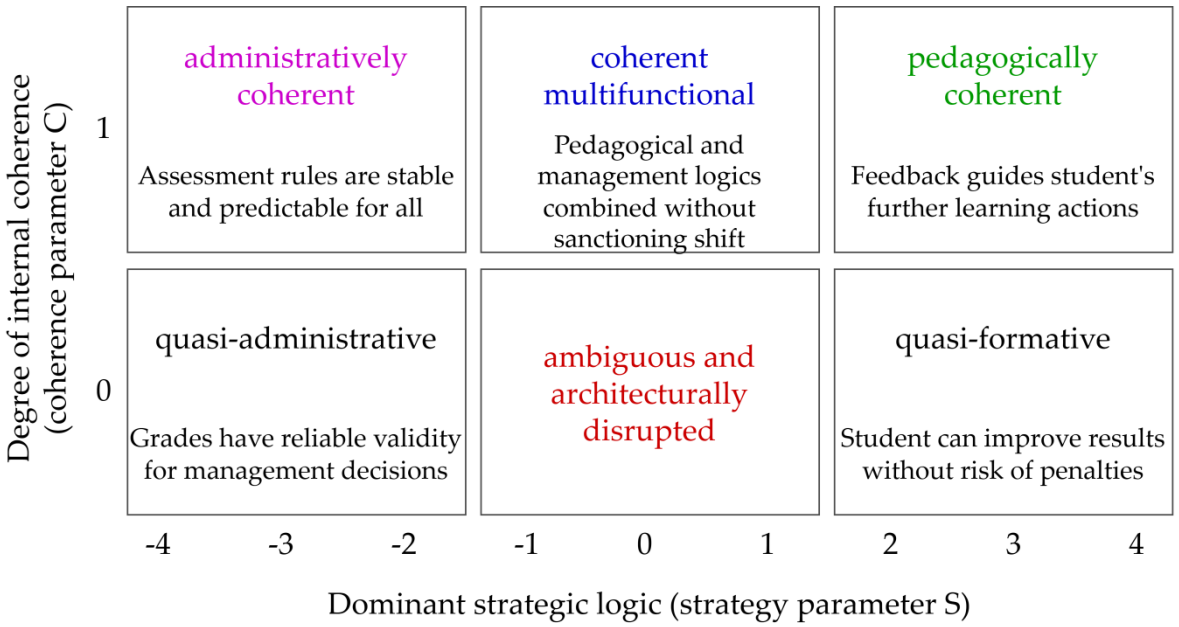

Fig. 2. Diagnostic Map of Assessment Architectures in Higher Education. Text color corresponds to trajectory color in Fig. 1.

Source: Created by authors

With a unit coherence parameter C, assessment systems can be divided by the value of strategy parameter S into:

- pedagogically coherent ($S \geq +2$, $C = 1$);
- administratively coherent ($S \leq -2$, $C = 1$);
- coherent multifunctional ($-2 < S < 2$, $C = 1$), strategically neutral but productive, combining pedagogical and administrative goals without internal contradictions.

Zero coherence parameter C indicates an architectural gap in three fundamentally different forms:

- quasi-formative trajectories ($S \geq +2$, $C = 0$), when pedagogical rhetoric masks control mechanisms;
- quasi-administrative trajectories ($S \leq -2$, $C = 0$), when the system loses effectiveness even as a control tool;
- ambiguous and architecturally disrupted ($-2 < S < 2$, $C = 0$), strategically neutral and unproductive, when different system levels implement contradictory logics.

4.4. Applicability of the Model at Different Levels

Before considering application of the model at institutional and academic levels, it is necessary to clarify the boundaries of the very concept of "assessment system" in interaction with external context.

In the proposed architecture, a fundamental distinction is made between:

- the assessment system as a pedagogical mechanism for forming and interpreting learning achievements (from goals and principles to grade formation mechanisms and regulatory parameters);
- secondary management use of assessment results, as practices that use a grade as an input indicator for management decisions (scholarship rankings, competitive selection, resource allocation).

Secondary use of results does not belong to the assessment system in the narrow sense, since it does not change the mechanism of final grade formation. However, the existence of such practices creates external pressure on pedagogical logic and must be taken into account when designing assessment systems. For example, an academic assesses a course work by a criterion-referenced rubric, a student receives 85 points – this is within the assessment system. The dean's office collects average grades and forms a ranking for scholarships – this is secondary use and is not an element of the assessment system.

The question arises: can the "green trajectory" (pedagogically coherent system) exist in a university where there are scholarship rankings? Answer: yes, if administrative use of results does not deform the internal mechanism of their formation. An architectural gap arises not from the very fact of external rankings, but from their interference in the assessment procedure (the grade depends on position in the group, not on work quality) or from deformation of pedagogical logic by regulatory parameters under pressure of external management requirements (as on the red trajectory).

Thus, critical from the point of view of pedagogical expediency are not rankings as such, but cases when ranking-based or sanctioning mechanisms are directly embedded in the procedure of assessing learning outcomes, creating internal contradiction between declared goals and actual assessment logic.

The proposed model is intended primarily for higher education institution administration, which develops or reviews institutional assessment policies. The model allows:

- analyzing existing assessment regulations for internal coherence;
- designing assessment systems for educational programs taking into account their specificity;
- diagnosing sources of conflicts between academics and students related to assessment;
- justifying decisions about changes in the assessment system through identification of architectural gaps.

Practical recommendations for different zones of the diagnostic map. Quasi-formative zone ($S \geq +2$, $C = 0$): the system declares pedagogical goals, but actually performs administrative control – a review of the regulatory parameters is required. Most common problems: excessive weight of summative control (over 40 %), strict exam admission thresholds, grade limitations on retakes, penalties for retaking. Solution: reduce weight of final control, replace sanctions with opportunities to improve results. Quasi-administrative zone ($S \leq -2$, $C = 0$): the system loses effectiveness even as a control tool. It is necessary to review coherence between administrative goals and the way results are used. Is ranking really needed if it is not used for selection? Strategically neutral unproductive zone ($-2 \leq S \leq 2$, $C = 0$): different system levels implement contradictory logics. One should analyze the sequence of decisions from goals to use of results to determine where the gap arises.

The architectural framework can also be used by an academic to design an assessment system in a syllabus. At the level of an individual discipline, an academic independently determines:

- assessment goals (supporting learning or controlling results);
- principles (what to compare: student progress or learning outcomes);
- representation form and grade formation mechanism;
- regulatory parameters (weights, thresholds, penalties).

An academic typically does not determine how results will be used at the institutional level (rankings, scholarships, selection). However, they can take this influence into account when designing their own system – for example, avoid excessive regulatory parameters if they know that results will be used for competitive selection.

Example of diagnosis of a typical assessment system from a syllabus of an academic discipline of a Ukrainian higher education institution:

- Criterion 1 (goal): "formative assessment" is declared in the syllabus → pedagogical feature ($S + 1$);
- Criterion 2 (temporal logic): cumulative system is used (60 % current + 40 % exam) → pedagogical feature ($S + 1$);
- Criterion 3 (regulatory parameters): there are strict restrictions (minimum 36 points for current work to be admitted to exam, maximum 75 points on retake, penalties for absences), which transform the system from supporting learning to controlling compliance with requirements → administrative feature ($S - 1$);
- Criterion 4 (form): summative assessment at course level → administrative feature ($S - 1$);
- Criterion 5 (coherence): declared formative goals contradict actual logic of sanctioning control through regulatory parameters → architectural gap ($C = 0$).

Conclusion: the system is strategically neutral (has two pedagogical and two administrative features) with an architectural gap between declared formative goals and actual implementation through sanctioning regulatory parameters. Thus, the system declares formative goals, but actually works as control-sanctioning. This explains typical conflicts: students perceive the system as unpredictable, academics complain about appeals, and administration does not understand the source of the problem.

5. CONCLUSIONS

The proposed architectural model allows analyzing assessment systems in higher education as a multilevel configuration of interrelated decisions – from conceptual goals to procedural mechanisms, regulatory parameters and ways of using results. Unlike existing typologies that classify systems by individual features (type of scale, form of control), the proposed approach reveals the actual function of assessment through analysis of coherence between architectural levels.

The key conclusion is that the problematic nature of a system is determined not by its strategic vector (pedagogical or administrative), but by internal architectural misalignment. Precisely gaps between declared goals, procedural mechanisms and regulatory parameters generate quasi-formative configurations (pedagogical rhetoric masks control mechanisms) and quasi-administrative ones (management logic loses transparency). A coherent system, on the contrary, can be effective in both pedagogical (supporting learning) and administrative (ensuring standards) scenarios.

The developed diagnostic tools allow quantitatively evaluating the strategic orientation of the system (parameter S from pedagogical to administrative) and its internal coherence (parameter C, indicating the presence of architectural gaps). Typical design trajectories (pedagogically coherent, multifunctional, quasi-formative, administratively coherent) illustrate how different combinations of decisions at architectural levels lead to fundamentally different educational practices and student behavioral strategies.

The proposed model implements the principle of balance between "desirable, possible and allowable" in higher education (John Biggs, 1996a), operationalizing it as a tool for designing assessment systems. The practical value of the model lies in the possibility of analytically distinguishing pedagogically meaningful configurations from systems with architectural gaps, which allows institutions to review their own practices from the position of actual function. Any assessment system is a compromise between pedagogical goals, technical constraints and administrative requirements; the architectural approach allows designing such compromises, minimizing the risk of architectural gaps and preserving conditions for mindful learning even under external pressure.

The model is intended for application at different organizational levels. At the institutional level, it allows designing and reviewing assessment policies, diagnosing sources of conflicts between academics and students, justifying changes through identification of architectural gaps. At the academic level, the model can be used for designing assessment systems in syllabi of individual disciplines, particularly for aligning goals, result representation forms and regulatory parameters taking into account institutional context.

Important is the distinction between the assessment system itself as a pedagogical mechanism and secondary management use of its results. Practices of forming rankings for scholarships or selection are not an element of the assessment system, but their existence must be taken into account when designing to avoid architectural gaps.

The prospect of further research is empirical testing of the diagnostic tool on a wide sample of institutional contexts, development of methodological recommendations for reviewing assessment systems, and analysis of long-term effects of different architectural configurations on educational outcomes and student behavior.

Author Contributions:
Conceptualization: Tetiana Zaporozhets.
Data curation: Anton Zaporozhets.
Formal analysis: Anton Zaporozhets, Tetiana Zaporozhets.
Investigation: Anton Zaporozhets, Tetiana Zaporozhets.
Methodology: Tetiana Zaporozhets, Anton Zaporozhets.
Supervision: Tetiana Zaporozhets.
Visualization: Anton Zaporozhets.
Writing – original draft: Anton Zaporozhets.
Writing – review & editing: Tetiana Zaporozhets, Anton Zaporozhets.

Acknowledgment: This research did not receive any outside support, including financial support.

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

[1] Andriichuk, O. (2023). Formative assessment as a tool for measuring the quality of the educational process: from theory to practice. *Youth & market*, 5(213), 149–154. https://doi.org/10.24919/2308-4634.2023.282775 (in Ukrainian)

[2] Anne Mesny, A., Roberge-Maltais, I., & Galy, A. (2026). Innovative assessment and grading practices in higher education: a critical exploration for management educators. *The International Journal of Management Education*, 24(1):101307. https://doi.org/10.1016/j.ijme.2025.101307

[3] Biggs, J. (1996a). Assessing Learning Quality: reconciling institutional, staff and educational demands. *Assessment & Evaluation in Higher Education*, 21(1), 5–16. https://doi.org/10.1080/0260293960210101

[4] Biggs, J. (1996b). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. https://doi.org/10.1007/bf00138871

[5] Boud, D., & Dawson, P. (2023). What feedback literate teachers do: an empirically-derived competency framework. *Assessment & Evaluation in Higher Education*, 48(2), 158–171. https://doi.org/10.1080/02602938.2021.1910928

[6] Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413. https://doi.org/10.1080/02602930600679050

[7] Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A Century of Grading Research. *Review of Educational Research*, 86(4), 803–848. https://doi.org/10.3102/0034654316672069

[8] Durm, M. W. (1993). An A is not an A is not an A: A History of Grading. *The Educational Forum*, 57(3), 294–297. https://doi.org/10.1080/00131729309335429

[9] European Commission: Directorate-General for Education, Youth, Sport and Culture. (2015). *ECTS users' guide 2015.* Publications Office of the European Union. URL: https://data.europa.eu/doi/10.2766/87192

[10] The Evergreen State College. (n.d.). *Narrative evaluations*. URL: https://www.evergreen.edu/our-learning-approach/narrative-evaluations

[11] Fanshawe College. (2022). *Prior learning assessment and recognition (PLAR) at Fanshawe College.* Fanshawe College Pressbooks. URL: https://ecampusontario.pressbooks.pub/plarinfofanshawestudents/

[12] Fischer, J., Bearman, M., Boud, D., & Tai, J. (2024). How does assessment drive learning? A focus on students' development of evaluative judgement. *Assessment & Evaluation in Higher Education*, 49(2), 233–245. https://doi.org/10.1080/02602938.2023.2206986

[13] Hampshire College. (n.d.). *Guide to the narrative evaluation system at Hampshire College*. URL: https://surli.cc/kssein

[14] Hoskin, K. (1979). The Examination, Disciplinary Power and Rational Schooling. *History of Education*, 8(2), 135–146. https://doi.org/10.1080/0046760790080205

[15] Howitz, W. J., McKnelly, K. J., & Link, R. D. (2025). Delving into the Design and Implementation of Specifications Grading Systems in Higher Education. *Education Sciences*, 15(1), 83. https://doi.org/10.3390/educsci15010083

[16] Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice,* 27(5), 480–500. https://doi.org/10.1080/0969594X.2020.1799190
[17] Mehvar, R. (2025). A Practical Guide to Grade Adjustment or Curving for Pharmacy and Other Professional Health Programs. *Pharmacy,* 13(1), 4. https://doi.org/10.3390/pharmacy13010004
[18] Minerva Project. (n.d.). *Assessment*. URL: https://www.minervaproject.com/assessment
[19] New College of Florida. (n.d.). *Explanation of the Grading Policy*. URL: https://catalog.ncf.edu/undergraduate/grading-policy
[20] Nilson, L. B., & Packowski, J. A. (2025). *Specifications Grading 2.0: Restoring Rigor, Motivating Students, Saving Faculty Time, and Developing Career Competencies* (2nd ed.). Routledge. https://doi.org/10.4324/9781003559658
[21] Remesal, A., & Estrada, F. G. (2023). Synchronous self-assessment: First experience for higher education instructors. *Front. Educ.* 8:1115259. https://doi.org/10.3389/feduc.2023.1115259
[22] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science,* 18(2), 119–144. https://doi.org/10.1007/bf00117714
[23] Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education,* 34(2), 159–179. https://doi.org/10.1080/02602930801956059
[24] University College Dublin. (n.d.). *Micro-credentials*. URL: https://www.ucd.ie/microcredentials/
[25] Wiliam, D. (2017). Assessment for learning: meeting the challenge of implementation. *Assessment in Education: Principles, Policy & Practice,* 25(6), 682–685. https://doi.org/10.1080/0969594x.2017.1401526
[26] Yin, S., Chen, F., & Chang, H. (2022). Assessment as Learning: How Does Peer Assessment Function in Students' Learning? *Front. Psychol.* 13:912568. https://doi.org/10.3389/fpsyg.2022.912568

Anton Zaporozhets, PhD student in the Department of Automation and Computer-Integrated Technologies, Bohdan Khmelnytsky National University of Cherkasy, Cherkasy, Ukraine;
ORCID ID: 0009-0002-4919-2157
Address: Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Boulevard, Cherkasy, 18031, Ukraine.
E-mail: ant.zaporozhets@gmail.com

Tetiana Zaporozhets, Doctor of Sciences in Physics and Mathematics, Professor, Professor in the Department of Physics, Bohdan Khmelnytsky National University of Cherkasy, Cherkasy, Ukraine;
ORCID ID: 0000-0002-1489-0441
Address: Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Boulevard, Cherkasy, 18031 Ukraine.
E-mail: zaptet@ukr.net

Received: January 08, 2026; **revised**: February 28, 2026; **accepted**: March 26, 2026; **published**: March 30, 2026.

Запорожець Антон, Запорожець Тетяна. Архітектура систем оцінювання у вищій освіті: багаторівнева модель і діагностичні критерії. *Журнал Прикарпатського університету імені Василя Стефаника*, **13** (1) (2026), 145-160.

Системи оцінювання у вищій освіті зазвичай описують через шкали, бали та формальні процедури. Проте чи можна, спираючись лише на тип шкали або інструменти, зрозуміти, яку функцію реально виконує оцінювання – підтримку навчання чи адміністративний контроль? Чому формально подібні системи в різних університетах призводять до принципово різних освітніх практик і поведінкових стратегій здобувачів? Оцінювання не може бути адекватно проаналізоване через окремі елементи, оскільки однакові за формою рішення можуть виконувати принципово різні функції залежно від механізму їх формування, поєднання та використання. Метою статті є розробка архітектурної аналітичної рамки для проєктування та діагностики систем оцінювання у вищій освіті як багаторівневої конфігурації взаємопов'язаних рішень. Для досягнення цієї мети систематизовано різнорідні характеристики систем оцінювання як варіанти рішень на

окремих рівнях архітектури, змодельовано типові конфігурації систем для різних інституційних цілей та розроблено діагностичні критерії для аналізу конструктивної узгодженості конкретної системи оцінювання. Наукова новизна полягає у запропонованому архітектурному підході, який розглядає оцінювання як багаторівневу конфігурацію рішень: концептуальний рівень визначає простір варіантів для методологічного рівня, який модифікується регуляторними параметрами та реалізується через організаційні форми. Представлено типові траєкторії оцінювання і діагностичний інструментарій для оцінки їх узгодженості. Модель операціоналізує принцип балансу між педагогічно бажаним, технічно можливим і адміністративно допустимим, щоб проєктувати системи для усвідомленого навчання навіть в умовах інституційних обмежень. Основний висновок: системи оцінювання є коректною лише за умови внутрішньої архітектурної узгодженості між рівнями.

Ключові слова: вища освіта, конструктивне узгодження, формувальне оцінювання, узгодженість оцінювання, дизайн оцінювання.