Computer Science and Interdisciplinary Research Journal
Volume 3, Issue 1, 2026. ISSN: 3033-1218, DOI: 10.70862/CSIR.2026.0301-21

Adaptive Education Platforms for Development of Algorithmic Thinking in Corporate Environment: Criteria for Selection of Experimental Environment

Tsvetelina Rashkova [1], Gencho Stoitsov [2]

[1,2] University of Plovdiv Paisii Hilendarski, 236 Bulgaria Blvd., 4003 Plovdiv, Bulgaria

[1] rashkova@uni-plovdiv.bg, [2] stoitzov@uni-plovdiv.bg

Abstract – This article develops a methodological framework for screening and subsequently verifying adaptive learning platforms intended to support algorithmic and critical thinking among adults in corporate settings. It combines a conceptual analysis of selected literature, document analysis, and multi-criteria modelling. Six criteria are operationalized through five-level descriptors, provisional normative weights, and non-compensatory thresholds. Documentation for five platforms is examined to distinguish vendor claims from demonstrated functionality and secured access rights. A separate numerical example using five hypothetical profiles illustrates weighted aggregation and sensitivity to weights and input assumptions; it is not an empirical ranking of the products. A schematic corporate case provides separate indicators for algorithmic thinking, critical thinking, and transfer. The main contribution is a two-stage procedure linking documentary screening to technical and expert verification. The framework requires validation. No intervention study is reported, and no platform-specific educational effect is established.

Keywords – adaptive learning, algorithmic thinking, critical thinking, corporate learning, multi-criteria evaluation, transfer of learning.

DOI: https://doi.org/10.70862/CSIR.2026.0301-21

Corresponding author: Gencho Stoitsov, University of Plovdiv Paisii Hilendarski
Email: stoitzov@uni-plovdiv.bg

Received: 4 October 2026.
Revised: 17 October 2026.
Accepted: 18 October 2026.
Published: 18 October 2026.

© 2026 Tsvetelina Rashkova, Gencho Stoitsov. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

1. Introduction

Corporate learning has practical value when employees can use acquired knowledge to solve new professional problems. Course completion, high satisfaction, and good performance on an immediate test describe different aspects of participation, but do not provide sufficient grounds for concluding that workplace behavior has changed. The classic model of Baldwin and Ford [1] treats transfer as dependent on learner characteristics, training design, and the work environment. Platform selection should therefore begin with the expected application of learning, rather than the number of available features.

Adaptive systems offer opportunities to modify the learning sequence and support in response to current performance. The meta-analysis by Wang et al. [2], covering 45 independent studies, found a positive average effect on cognitive learning outcomes compared with non-adaptive interventions ($g = 0.70$). However, effect size depends on the context, duration, and design of the studies. This finding supports interest in adaptivity, but does not justify attributing the average effect to a specific product or to the development of a particular form of thinking within an organization.

The reviews by Martin et al. [3] and Kabudi, Pappas, and Olsen [4] show considerable diversity in adaptive approaches, technological implementations, and measured outcomes. Bulgarian literature also addresses the pedagogical opportunities and limitations of artificial intelligence in education [5]. A significant challenge for corporate practice remains translating general technological characteristics into

verifiable requirements for tasks, feedback, and assessment.

This article addresses that challenge through a framework for selecting an environment for learning and evaluating its outcomes. Its central argument is that a high functional rating is insufficient: a platform may be useful for distributing content while offering no means of observing reasoning or independently investigating transfer. Pedagogical suitability and the strength of supporting evidence are therefore assessed together, and critical limitations cannot be offset by high scores on other indicators.

2. Aim and Analytical Questions

The article aims to develop and critically apply a transparent procedure for evaluating adaptive platforms against the requirements for developing and measuring algorithmic and critical thinking in corporate settings. The analysis examines the alignment between pedagogical objectives, technological mechanisms, and conditions for verifying outcomes. The platform itself is treated as a means of organizing and observing learning activities, rather than an independent explanation for changes in competence.

The analysis addresses three questions. First, which observable characteristics distinguish adaptive learning support from personalized content recommendations? Second, how can these characteristics be linked to separate indicators of algorithmic and critical thinking? Third, how can a selection decision remain defensible when product information is incomplete, access is conditional, and organizational priorities differ?

The study is theoretical and methodological, with an illustrative multi-criteria application. Its outcome is a model for informed decision-making and an analysis of its limitations. It does not test a causal hypothesis about the effectiveness of adaptive learning, because no primary data from an intervention involving learners are presented.

3. Theoretical and Conceptual Framework

3.1. Algorithmic and Critical Thinking

Algorithmic thinking involves formulating a problem, specifying inputs and constraints, constructing a precise sequence of actions, and checking the solution [6]. In corporate settings, it may be demonstrated when developing a procedure for handling requests, detecting errors in an information flow, or organizing a sequence of control activities. Learners need to justify the order of operations, branching conditions, and cases in which the procedure fails.

The concept is related to computational thinking but does not encompass it in full. Abstraction, decomposition, and problem representation belong to a broader set of problem-solving approaches [7,8]. Performance on a programming task should therefore not automatically be interpreted as a general measure of algorithmic thinking. The logic of the solution, syntactic skills, and prior technical preparation must be distinguished.

Critical thinking refers to purposeful, self-regulatory judgement involving analysis, evaluation of reasons, inference, and explanation of a decision [9]. In a professional situation, this means distinguishing facts from assumptions, reliable sources from unverified claims, and a valid procedure from the mechanical application of a rule. Halpern [10] links critical thinking instruction to deliberate support for transfer across situations.

The two constructs are complementary. An employee may flawlessly execute a sequence based on an incorrect assumption, or identify a weakness in an argument without developing an executable procedure. An overall score for “problem-solving” would therefore conceal important differences. The proposed framework requires separate indicators and allows the relationship between them to be analyzed after each construct has been assessed independently.

3.2. Adaptation and Pedagogical Mechanism

An adaptive learning environment uses information about the learner to modify content presentation, navigation, or support [11]. This article distinguishes resource recommendations, branching learning paths, and adaptation during problem-solving. Recommending a course based on a job role is a form of personalization, but does not in itself demonstrate that the system responds to a specific error or strategy.

The pedagogically meaningful relationship is between observation, interpretation, and subsequent action. For example, an incorrect sequence may trigger a task on dependencies between steps, whereas a correct solution accompanied by unjustifiably high confidence may be followed by a counterexample. These are design possibilities that must be demonstrated using specific content. The presence of artificial intelligence, automated recommendations, or a chat interface is insufficient evidence of these capabilities.

Response time and the number of attempts are useful process data, but permit different interpretations. A quick response may indicate mastery or guessing; repeated attempts may reflect persistence or misunderstanding. The learner model must therefore be compared with solution quality, the reasons provided, and performance on a new task. Self-reported confidence provides information about

judgement calibration, but does not replace the measurement of knowledge and thinking.

3.3. Adult Learners and Transfer

The andragogical perspective draws attention to professional experience, autonomy, and the practical relevance of tasks [12]. This does not mean that all adults learn in the same way. Differences in prior knowledge, digital skills, and confidence require diagnostic assessment and opportunities for support. Adaptation should support autonomy, including through understandable feedback and opportunities for learners to reconsider their own strategies.

Bloom's revised taxonomy helps distinguish recall from analysis, evaluation, and the creation of a solution [13]. ADDIE organizes the alignment of needs, instructional design, and evaluation [14]. These approaches serve different functions: the taxonomy specifies cognitive demands, while the instructional model structures development. Neither demonstrates effectiveness without evidence of performance and outcomes.

Transfer must be distinguished from solving a familiar task again. Near transfer may be assessed using a new case with the same logical structure; more distant transfer requires applying a principle to a different professional situation. Supervisor support, available time, and opportunities to apply learning remain important conditions. Consequently, a lack of observed change cannot be attributed solely to the platform in advance.

4. Methodological Approach and Evidence Base

The approach combines a conceptual synthesis of selected scientific sources, document analysis, and the construction of an evaluation procedure. Theoretical sources support the concepts and their relationships; review publications outline the available evidence; product sources identify declared functionality. These three types of information are not treated as interchangeable.

The literature analysis is targeted and does not claim to be exhaustive. A complete search log, a count of screened records, and independent duplicate screening are unavailable; the article is therefore not presented as a systematic review or a completed scoping review. PRISMA-ScR is a reporting standard for scoping reviews, rather than a substitute for conducting and documenting a search [15]. This distinction defines the limits of the generalizations supported by the present study.

The comparison set includes Coursera for Business, Docebo, Litmos, Area9 Rhapsode, and Cornerstone Learning. It is purposively limited and represents different combinations of corporate administration, personalization, and adaptive functionality. Market representativeness is not claimed. Considering learning management systems (LMS), content libraries, and specialized adaptive solutions within a common set is justified only in relation to a specific educational objective; it does not imply that the products are equivalent across all applications.

The document analysis is based on official product sources checked on 14 October 2026; the findings are organized in Appendix B. No technical tests were conducted, and a comparable licensing scope was not established. The numerical application therefore uses hypothetical profiles that are separate from the actual products. It demonstrates the procedure without constituting measurement or a market ranking.

The evidence register records the platform, criterion, source, date checked, specific function, coding rationale, and unconfirmed conditions. Appendix B provides such a register for the corpus examined. "D" denotes functionality declared in a public source; "T" denotes functionality technically demonstrated through a common test protocol; "P" denotes permissions confirmed in writing for a specific configuration; and "UC" denotes an unconfirmed descriptor. These statuses are not quality levels and may be combined. No "T" or "P" statuses are assigned in the present analysis. An incomplete description is not automatically coded as a score of 1 and does not demonstrate that a capability is absent.

5. Evaluation Framework

5.1. Criteria and Operationalization

The framework comprises six criteria. K1 assesses the degree and temporal scale of adaptation; K2 assesses capabilities for complex cognitive tasks; K3 covers corporate applicability; K4 assesses observability and the extraction of process data; K5 describes independent scientific documentation; and K6 covers practical access for verification and research. Full descriptors are provided in Appendix A.

The criteria distinguish technical capability from pedagogical outcomes. K2, for example, does not measure whether thinking has already developed; it assesses whether tasks that require and record reasoning can be implemented. The presence of an administrative dashboard does not satisfy K4 if individual attempts and the support received cannot be traced. K5 concerns evidence about the specific platform and its application, rather than general publications on adaptive learning.

To limit overlap, K4 concerns the technical structure and export of data, while K6 concerns permissions, licensing, and organizational arrangements for access. Published support for xAPI

integration may indicate technical potential under K4, but does not establish that the research team will receive all necessary events under a particular contract. Similarly, a high K3 score cannot compensate for a lack of tasks aligned with the target constructs.

Table 1 links the two constructs to examples of observable performance. The indicators are proposed for task design and require validation; they do not constitute a ready-to-use standardized test.

Table 1. Constructs and observable indicators

Component	Example task	Observable indicator
Decomposition	Break down a request-handling process	Complete steps and correct dependencies
Algorithmic correctness	Procedure with conditions and exceptions	Logical sequence and handled boundary cases
Checking and improvement	Identify a defect in a work procedure	Appropriate counterexample and justified correction
Critical evaluation	Compare conflicting sources	Distinguish facts, assumptions, and limitations
Reasoned choice	Choose between alternative solutions	Relevant reasons and consideration of counterarguments
Transfer	New case with changed conditions	Independent application and adaptation of the principle

5.2. Weights and Eligibility Requirements

The provisional normative weights are 0.25 for K1 and K2, 0.15 for K3 and K4, and 0.10 for K5 and K6. Thus, half of the weight is assigned to the adaptive mechanism and cognitive design, 0.30 to corporate applicability and observability, and 0.20 to the scientific evidence base and access. This allocation and the equal weights within each pair are the authors' design assumptions, rather than results of an expert survey or optimization. In practice, priorities should be established before evaluating products and examined through sensitivity analysis.

The overall score S_j is calculated as the sum of the products of weight w_i and rating x_{ij} across the six criteria. Ratings range from 1 to 5, and the weights sum to one. The weighted score supports selection; it is not a measure of educational effect. Summation assumes provisionally equal steps between categories on the ordinal scale. A difference of 0.20 points should therefore not be interpreted as a precise quantity of superior quality.

$$S_j = \sum_{i=1}^{6} w_i\, x_{ij};\ x_{ij} \in \{1, \ldots, 5\};\ \sum_{i=1}^{6} w_i = 1;\ w_i \geq 0$$

Here, j denotes the profile evaluated and i the criterion. A final S_j is not calculated if any criterion is UC. Renormalising the weights over the known criteria alone would change the decision problem and could conceal a critical gap; it is therefore not applied. All input values are explicitly specified in the numerical demonstration.

Final eligibility requires four non-compensatory conditions: K1 ≥ 4, K2 ≥ 4, K4 ≥ 3, and K6 ≥ 3. K1 requires adaptation within the session; K2 requires an adaptive, multistep case with alternative paths and recorded reasoning; K4 requires export at least at task level; and K6 requires access secured in writing. These thresholds are design requirements for the article's purpose and require expert validation. Failure to meet a condition cannot be offset by the overall score; an unknown condition results in unconfirmed eligibility.

The procedure has two stages. The first is documentary screening, which identifies potentially relevant functions, evidence gaps, and questions for demonstration. The second is technical and expert verification using the same case, a control export, and a license specified in writing. The final statuses are "eligible", "ineligible due to a demonstrably unmet requirement", and "eligibility unconfirmed". Final ranking is restricted to configurations with confirmed eligibility. Where evidence is incomplete, screening identifies the checks required but does not establish a winner.

The K2 ≥ 4 threshold is tested using the case in Appendix C. Demonstration must cover a multistep solution, differentiated support based on performance, retention of reasoning, and the possibility of independently assessing the two constructs. The latter may be undertaken outside the platform using exported responses; automated assessment of free text is not mandatory. Passing the threshold requires documented fulfilment of all conditions, rather than merely the presence of tests or scenarios in a catalogue.

5.3. Reliability of Assessment

Reliable application requires at least two assessors to use the same evidence corpus and assign scores independently. Disagreements should be analyzed by criterion before an overall assessment is formed. Agreement for ordinal data should be reported, for example using weighted kappa for two assessors, with the weighting scheme and category distribution specified. No such coefficients are calculated in this application because independent assessment records are unavailable.

The framework's content validity requires review by specialists in adult learning, educational technology, and measurement. Expert assessment

should establish whether the descriptors are clear, observable, and sufficient to distinguish adjacent levels. A multi-criteria procedure does not in itself guarantee validity. It provides traceability that allows assumptions to be discussed and revised.

The proposed validation sequence comprises expert content review, independent coding of the same evidence, pilot technical verification, and analysis of the relationship between the criterion profile and the tasks actually implemented. Only then is an empirical effectiveness study justified. These activities are future steps; the article reports no findings from them.

6. Illustrative Application and Sensitivity Analysis

6.1. Documentary Screening and Hypothetical Numerical Profiles

Table 2 specifies five hypothetical profiles, A–E, to demonstrate aggregation. Profiles B and E intentionally have identical values, illustrating the possibility of a tie. The numerical assumptions are not assigned to actual platforms. Available product evidence is examined separately in Appendix B.

Table 2. Hypothetical input profiles used to demonstrate the method

Hypothetical profile	K1	K2	K3	K4	K5	K6	Score
Profile A	2	3	5	3	3	2	2.95
Profile B	3	3	5	4	2	2	3.25
Profile C	2	3	5	3	2	3	2.95
Profile D	5	4	4	5	3	3	4.20
Profile E	3	3	5	4	2	2	3.25

Note. The profiles are hypothetical, rather than anonymized empirical results. Unconfirmed values for actual platforms are marked UC, and no final aggregate is calculated for them.

The documentation for Area9 Rhapsode describes the use of performance and self-reported confidence, as well as the transmission of xAPI statements to a learning record store [16,17]. This provides specific grounds for questions under K1 and K4, but does not confirm all descriptor conditions. Support for xAPI does not establish verified completeness and sequencing of exports for K4 = 5. The description of flexible licensing does not constitute access secured in writing for K6 = 3 [18]. Independent technical and contractual confirmation is needed for these ratings.

Honke and Becker-Genschow [19] examine the use of an adaptive environment in bionics instruction involving 56 sixth-grade pupils. The study is relevant as an example of platform use and learning outcome measurement. Differences in age, content, and organizational conditions limit the transfer of conclusions to employees. Improvement in subject knowledge does not in itself establish the development of algorithmic or critical thinking.

Product sources for Coursera for Business, Docebo, Litmos, and Cornerstone Learning describe corporate learning and personalization [20-23]. They do not provide a comparable, complete set of ratings across K1–K6. For example, a personalized path does not automatically establish adaptation following a specific error, and administrative reporting does not guarantee the export of individual attempts. Appendix B distinguishes documented claims from missing grounds for numerical coding.

6.2. Sensitivity Towards Weights

Four scenarios were calculated with the hypothetical inputs held constant: baseline, equal weights, priority to cognitive tasks, and priority to corporate implementation. The latter two are analytical assumptions, rather than established preferences of an expert panel. The vectors are ordered from K1 to K6: baseline (0.25, 0.25, 0.15, 0.15, 0.10, 0.10); equal (1/6 each); cognitive (0.25, 0.35, 0.10, 0.15, 0.10, 0.05); and corporate (0.15, 0.15, 0.35, 0.15, 0.10, 0.10).

Table 3. Weighted scores of hypothetical profiles under four scenarios

Hypothetical profile	Baseline	Equal	Cognitive	Corporate
Profile A	2.95	3.00	2.90	3.45
Profile B	3.25	3.17	3.20	3.65
Profile C	2.95	3.00	2.85	3.45.
Profile D	4.20	4.00	4.25	4.10
Profile E	3.25	3.17	3.20	3.65

Profile D has the highest score in every scenario. Its lead over the tied profiles B and E is 0.95, 0.83, 1.05, and 0.45, respectively. Only D passes all four thresholds in the hypothetical example; the other profiles have K1 < 4 and K2 < 4. Table 3 therefore compares mathematical aggregates for all profiles, rather than providing a final ranking of eligible solutions. Values are rounded to two decimal places after the full sums have been calculated.

If the weight of K3 varies while the other weights retain their baseline relative proportions, profiles D and B are tied at $w_3 \approx 0.564$. Above this value, B has a higher aggregate, but remains ineligible under K1 and K2. A change in numerical ordering is therefore not necessarily a change in the final selection.

6.3. Sensitivity to Uncertainty in Ratings

For profile D, reducing K2 from 4 to 3, K4 from 5 to 3, and K6 from 3 to 2 lowers the baseline score from 4.20 to 3.55. The aggregate remains above the 3.25 obtained by B and E, but the K2 and K6 thresholds are no longer met. In this scenario, no profile passes all conditions. This is a valid outcome of the procedure: the configuration must be reconsidered or the

candidate set expanded, rather than automatically selecting the highest aggregate. The reduced values are assumptions; where data are absent, UC is recorded.

Similarly, reducing K5 from 3 to 1 for hypothetical profile A changes its baseline score from 2.95 to 2.75. This checks the arithmetic effect of an assumption without making any claim about the presence or absence of publications on a particular product. If intervals are used for unconfirmed ratings, their bounds must be supported by evidence. For a completely unknown rating, the permissible range is the entire 1–5 scale. The scenarios are deterministic and have no probabilistic interpretation.

Documentary screening concludes with the verification programme in Appendix B. No actual platform is certified as finally eligible within the corpus examined. The next step is technical and expert testing using a common case and the four conditions in Section 5.2.

7. Pedagogical and Measurement Requirements for Implementation

Appendix C presents a common case for technical verification: a customer request involving incomplete data and conflicting instructions. It includes the expected logic, support rules, tests, and separate rubrics for the two forms of thinking. Implementing it in each configuration serves to verify K1, K2, and K4.

Algorithmic thinking can be assessed through completeness, logical correctness, exception handling, and verification using test cases. Critical thinking is assessed through the reliability of reasons, consideration of alternatives, and justification of limitations. Analytical rubrics must distinguish a correct structure with incomplete argumentation from a persuasively presented but logically unworkable solution. Assessment tasks should not be identical to practice tasks.

The minimum set of process data may include a pseudonymous identifier, task identifier and version, attempt number, response, support received, time, and an available code for the adaptive action. The reason for selecting the next task is recorded when the platform provides it. If it is absent, the analysis must acknowledge the limited observability of the mechanism. An exchange standard is not assumed to guarantee all these fields automatically.

Effectiveness evaluation requires baseline, immediate post-test, and delayed measurements using equivalent but different tasks. The interval before the delayed assessment should be specified in advance according to the work cycle. Independent assessment of open-ended solutions, where possible without knowledge of the participant's group, reduces the risk that expectations about the technology will influence scoring.

Comparing adaptive and non-adaptive interventions requires a clear definition of the effect being investigated. With equal duration, differences in attained proficiency can be assessed; with the same required mastery level, the time needed can be compared. If content, instructor support, and learning exposure differ simultaneously, the outcome cannot be attributed solely to adaptation. Randomization, or transparently described group matching where randomization is impracticable, helps limit this problem.

Corporate settings also require attention to dependence among participants from the same team, differences in prior preparation, and attrition. Sample size should be justified in relation to the selected primary outcome and design, rather than an arbitrary fixed number of employees. Process data may explain different learning trajectories, but should not automatically substitute for final outcomes.

8. Discussion and Limitations

The proposed framework shifts attention from the general label "adaptive platform" to a verifiable relationship between technology and educational objectives. Its contribution lies not in using a weighted sum, but in combining criterion-based ordering with conditions that can prevent a final selection. The analysis shows that stable numerical leadership and implementation feasibility are different properties.

The practical advantage of the procedure is that contested assumptions can be located precisely. If evidence of task adaptation is missing, K1 must be checked; if reasoning data are unavailable, the issue concerns K2 and content design; if export rights are absent, the obstacle falls under K6. Selection can thus conclude with a specific programme of technical verification rather than an unsupported product recommendation.

The framework has substantial limitations. The targeted literature selection is not exhaustive, and the five platforms do not represent the market. Documentary evidence has not been technically verified and does not support a complete numerical comparison of products. The hypothetical profiles examine the internal logic of aggregation without validating the scales, weights, or thresholds. Product pages may change; practical application requires archived sources, a fixed version and package, and records of equivalent checks.

The scales are ordinal and partly composite. A high K2 level, for example, simultaneously requires a complex task, appropriate adaptation, and observation of reasoning. This creates a risk that two different functional profiles receive the same score. In practical

implementation, the overall assessment must be retained alongside the detailed profile, evidence, and assessors' disagreements.

Scientific documentation also requires careful interpretation. More publications may reflect a longer market presence or better access for researchers, rather than greater effectiveness. A study using a particular platform is relevant only insofar as its tasks, participants, and measures are comparable to the purpose of the selection. K5 is therefore used as an indicator of the available evidence base, rather than an automatic measure of quality.

Ethical applicability requires transparency about the behavioral data collected and the purposes of assessment. Research participation should include appropriate information, voluntariness, and separation of research results from routine managerial appraisal of employees. Proposed design measures include a limited dataset, pseudonymization, defined access roles, and a retention period agreed in advance. These are requirements for responsible design, rather than a report of an already approved ethics protocol.

9. Conclusion

Selecting an adaptive learning platform for developing algorithmic and critical thinking requires alignment at three levels: the reasoning expected of learners, the way the system modifies support, and the means of independently verifying outcomes. The article proposes a six-criterion framework that makes these relationships explicit and distinguishes functional assessment from confirmed eligibility.

The numerical application shows that a stable weighted aggregate does not guarantee eligibility: a change in cognitive capabilities or access may exclude the leading profile. Document analysis identifies which claims about actual platforms require subsequent verification. The framework thus links screening to a verifiable decision without attributing educational effects to the technology.

The methodological value of the framework lies in the transparency of selection: every decision can be traced to a criterion, assumption, and piece of evidence. Further validation requires independent assessors, verification in actual configurations, and linkage of process data to separate measures of the two constructs and transfer. For corporate practice, the most defensible choice is an environment that enables both the delivery of learning and critical examination of its outcomes.

Appendix A. Operational Descriptors of the Criteria

The highest level for which all stated requirements are met is assigned. Where evidence is insufficient, "unconfirmed" is recorded and a range may be analyzed if necessary. For K6, level 3 requires access to have actually been secured; a general description of flexible licensing is insufficient. The descriptors constitute a proposed rubric requiring expert review.

Table 4. K1 Degree of adaptivity

Score	Descriptor
1	Identical static content and sequence for all participants.
2	Recommendation of a course or resource before the learning activity begins.
3	Branching after a test or personalization at learning-path level.
4	Modification of a module, sequence, or difficulty within the session based on performance.
5	Task-level adaptation using an updated learner model and more than one signal, such as accuracy and confidence.

Table 5. K2 Capabilities for cognitive tasks

Score	Descriptor
1	Closed-ended tasks primarily focused on recalling information.
2	Tasks requiring understanding and standard application, without multistep scenarios.
3	Analysis, scenarios, or projects without demonstrated adaptation to the problem-solving strategy.
4	Adaptive multistep tasks, alternative paths, and recorded reasoning.
5	Adaptation of tasks and feedback to the type of error or reasoning, with traceable assessment indicators.

Table 6. K3 Corporate applicability

Score	Descriptor
1	No functions for organizational roles and groups
2	Limited pilot use with partial administrative functions
3	Support for groups, roles, and basic integrations with learning systems.
4	Documented corporate implementations and integrations through SSO or APIs.
5	Support for multiple audiences, large-scale administration, and documented security mechanisms.

Table 7. K4 Observability and data export

Score	Descriptor
1	Completion information only, without usable export.
2	Summary reports without data on individual actions.
3	Export of results at task or interaction level.
4	Export or API with timestamps and identifiers linking participants, tasks, and groups.
5	Event data through xAPI/LRS or an equivalent mechanism, with verified completeness, sequencing, and traceability.

Table 8. K5 Independent scientific documentation

Score	Descriptor
1	Only marketing claims within the corpus examined
2	Vendor case studies or non-peer-reviewed evaluations.
3	At least one external peer-reviewed study of the specific platform, with clearly described limitations.
4	Several independent peer-reviewed studies in actual learning environments.
5	Independent replications across contexts, including adult or corporate learning.

Table 9. K6 Practical access for verification

Score	Descriptor
1	Established lack of authorized access for the required verification or research.
2	Demonstration access without confirmed rights for groups and research data.
3	A pilot or flexible license secured in writing, with the necessary groups and data rights for the specified period.
4	A research or equivalent license with confirmed administrator rights and technical assistance.
5	A formal partnership guaranteeing groups, data, technical support, and conditions for replication.

Appendix B. Evidence register for documentary screening

The register covers the listed official sources, checked on 14 October 2026, and the cited scientific publication. Product documents describe claimed capabilities; a specific version, contract, and test protocol have not been established. "UC" refers to the sufficiency of evidence for the full descriptor, rather than the absence of a function. No final aggregate score is derived for actual platforms.

Coursera for Business

Source: Coursera [21], Online business learning platform.

K1. D: personalized learning paths. UC: the timing and rules of adaptation; no exact score is assigned.

K2. D: practical formats and projects. UC: implementation of a common multistep case and recording of reasoning.

K3. D: corporate learning and organizational clients. UC: the full descriptor for the specific package.

K4. UC: export of attempts, timestamps, and linking identifiers has not been verified.

K5. UC: the cited corpus provides no basis for a platform-specific scientific documentation score; this does not imply that studies are absent.

K6. UC: contractual access has not been presented; an invitation to make contact does not establish an actual demonstration or pilot license.

Docebo

Source: Docebo [22], official product page.

K1. D: personalized content and paths. UC: response to a specific attempt in the common case.

K2. D: interactive content and scenarios. UC: recording of and adaptation to reasoning.

K3. D: multiple audiences, integrations, and security information. A documentary code of 5 is possible; applicability to a specific package remains UC.

K4. D: reporting and dashboards. UC: complete export of process fields; administrative indicators do not justify a score of 4.

K5. D: vendor customer stories corresponding to the evidence type for level 2. They do not establish independent effectiveness or an exhaustive scientific corpus.

K6. UC: an offer of a demonstration does not prove that access and data rights have been granted.

Litmos

Source: Litmos [23], Litmos AI.

K1. D: recommendations based on role and learning history, corresponding to level 2. A higher level remains UC.

K2. D: content and question generation. UC: implementation of the multistep case and traceability of argumentation.

K3. D: automated enrolment and learning processes. UC: fulfilment of all conditions of descriptor 5 for the specific configuration.

K4. D: automated reporting. UC: granularity and completeness of attempt exports.

K5. D: customer testimonials corresponding to the evidence type for level 2; these are not independent peer-reviewed studies.

K6. UC: a pilot and data rights have not been secured in writing; a score of 3 cannot be confirmed.

Area9 Rhapsode

Sources: Area9 Lyceum [16-18]; Honke and Becker-Genschow [19].

K1. D: performance and confidence as signals. UC: technical demonstration of all conditions for level 5.

K2. UC: the cited information does not demonstrate an adaptive, multistep corporate case with recorded reasoning; a score of 4 is not substantiated.

K3. UC: the help pages used do not substantiate the full corporate descriptor for level 4.

K4. D: transmission of xAPI statements to an LRS. UC: scope and verified completeness; a score of 5 is not substantiated.

K5. The 2025 publication describes platform use with school pupils and supports the presence of an external peer-reviewed study for level 3. It does not demonstrate a corporate effect or independent replications. The authors acknowledge vendor support for the modules and implementation, which limits interpretation as a fully independent evaluation.

K6. D: different licensing models. UC: an actually secured pilot license and data rights; a score of 3 is not substantiated.

Cornerstone Learning

Source: Cornerstone [20], Learning management system.

K1. D: personalization based on learning behavior and skills. UC: specific rules and temporal scale within the selected module.

K2. D: interactive learning and specific role-play capabilities. UC: their inclusion in the same package and recording of reasoning.

K3. D: corporate learning management. UC: all level 5 requirements within the specific scope.

K4. D: tracking and reporting. UC: export of linked events with timestamps for the common case.

K5. D: customer testimonials corresponding to the evidence type for level 2; scientific independence cannot be inferred from them.

K6. UC: contractual rights for a pilot, groups, and research data export have not been established.

Protocol for the Next Stage

For each platform, the package, version, and date are fixed, the same test cases are executed, and the expected log is compared with the export obtained. Two assessors independently code each criterion and record the grounds for assigning the highest fully satisfied descriptor. For K5, a separate search using the platform name and previous names is documented, with checks of peer review, funding, and context. For K6, written confirmation of groups, duration, roles, and data rights is attached. Final eligibility is not declared when a threshold remains unconfirmed.

Appendix C. Schematic case for handling a customer request

C1 Purpose and initial conditions

The case is an original instructional example using hypothetical internal rules. It supports task design and technical demonstration; it does not describe a real organization, a completed intervention, or a legal requirement. Its purpose is to distinguish the construction of an executable procedure from critical evaluation of the reasons for a decision.

An employee receives a request to correct a charge of 120 notional units. The request number and the customer's identity have been verified, but a statement explaining the basis for the charge is missing. The customer demands an immediate decision. The repository contains the current instruction R2 and an archived instruction R1; a colleague has sent a message saying "process immediately". The deadline for an initial response is two working hours and is not a deadline for final approval.

R2 requires the following: if the data are incomplete, request the missing statement and record a pending status; if the data are complete and an erroneous charge of up to and including 100 units is established, the employee may approve the correction; amounts above 100 units must be referred to a supervisor; if the charge is valid, send a reasoned refusal. Archived R1 allows independent approval up to 150 units. Current R2 takes precedence over R1 and the informal message.

Learner task: identify the verified facts and unknowns; present an algorithm as steps or pseudocode; justify the choice of rule; propose boundary tests; and formulate an initial response to the customer. After receiving a statement confirming the erroneous charge, update the decision.

C2 Expected logic and test cases

The expected sequence is: check completeness → if the statement is missing, request information, notify the customer, and set the status to "pending" → once the statement is received, check the basis of the charge → if the charge is valid, issue a reasoned refusal → if it is erroneous, check the threshold → approve amounts up to and including 100 units and refer amounts above 100 to a supervisor. Each decision is recorded with the version of the rule applied. Approval is not justified on the initial information; after the error is confirmed, the amount of 120 requires escalation.

Minimum tests: incomplete data and 120 → pending; complete data, an error, and 100 → independent approval; complete data, an error, and 101 → escalation; complete data, no error, and 80 → reasoned refusal. Different representations of the algorithm are acceptable if they preserve these dependencies and do not replace missing information with assumptions.

C3 Adaptive support and observability

If the learner approves a correction despite missing data, a prompt about prerequisites is provided, followed by another attempt. If R1 is used, the learner is invited to compare the status of the two documents. If the boundary operator is incorrect, the cases of 100 and 101 are presented. If the decision is correct but the justification incomplete, the learner is asked to explain why the informal message does not override R2. A correct and justified decision is followed by a new variant without a prompt. These are predefined instructional design rules; it is not assumed that all platforms can execute them automatically.

For each attempt, checks cover the recording of a pseudonymous identifier, case version, response, time, error type, displayed prompt, and next task. Self-reported confidence is recorded separately, if available, and earns no correctness points. Comparing expected and exported events tests K4; the ability to provide different support according to the error tests K1 and K2.

C4 Separate assessment and transfer

The rubrics in Tables 10 and 11 each contain four indicators scored 0, 1, or 2. The two results are reported separately, each ranging from 0 to 8; they are not added into an overall thinking score. The first independent solution without prompts is assessed.

Performance after support is stored separately to avoid conflating independent skill with assisted performance.

Each assessor selects the highest level whose conditions are fully met and cites an excerpt or step from the response as evidence. A response not provided for an indicator receives 0; a technically lost or unpresented response is recorded as a missing measurement, and no total is calculated for the relevant construct. Two assessors work independently and then discuss discrepancies. Before use, the rubrics require pilot calibration with sample responses. No validated thresholds for successful learning have been established.

Table 10. Analytical rubric for algorithmic thinking

Indicator	0 points	1 point	2 points
A1 Input checks	Does not check data sufficiency and permits a decision without the statement.	Recognizes the missing statement but specifies neither an action nor a condition for proceeding.	Checks the identifier, identity, and statement; requests missing information and proceeds after receiving it.
A2 Procedural logic	No executable procedure is provided, or the main decisions are confused.	Distinguishes at least two main outcomes but omits a branch or dependency.	Orders the checks and distinguishes pending status, refusal, approval, and escalation according to the stated conditions.
A3 Boundary cases	Both boundary cases are handled incorrectly, or no solution is provided for them.	Handles only one of the cases 100 and 101 correctly, or lists them without a general condition.	Uses ≤ 100 for approval and > 100 for escalation when data are complete and an error is established.
A4 Verification through tests	Provides no test with inputs and an expected output.	Proposes one to three of the four tests in C2 with correct expected outputs.	Covers all four tests in C2 with correct inputs and outputs and shows how the procedure handles them.

Table 11. Analytical rubric for critical thinking

Indicator	0 points	1 point	2 points
CT1 Facts and assumptions	Presents the customer's claim of an erroneous charge as an established fact.	Identifies missing evidence but does not clearly distinguish known data from claims.	Separates verified data, the claim of an error, and the missing statement; justifies the need for verification.
CT2 Source evaluation	Bases the decision on R1 or the message without considering current R2.	Selects R2 but incompletely explains why the other two sources do not override it.	Justifies the priority of current R2, the archived status of R1, and the lack of authority of the informal instruction.
CT3 Comparing alternatives	States a decision without considering another possibility.	Compares two possibilities but omits a relevant condition or consequence.	Compares immediate approval, requesting data, and subsequent escalation; explains when each option is justified or inadmissible.
CT4 Limits of the conclusion	Makes a final commitment without evidence or beyond their authority.	Identifies a limitation but does not explain how new data would change the decision.	Formulates a conditional conclusion, accounts for authority, and distinguishes an initial response from final approval.

Near transfer is assessed using a new request with different amounts and a changed current threshold, without prompts. More distant transfer may be assessed through an internal access request with an analogous relationship between data sufficiency, authority, and escalation. The new context is not assumed in advance to be equivalent in difficulty. Transfer results are reported separately from practice results and interpreted after task comparability has been checked.

References

[1] Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*(1), 63–105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

[2] Wang, X., Huang, R. T., Sommer, M., Pei, B., et al. (2024). The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis. *Journal of Educational Computing Research*, *62*(6), 1348–1383. https://doi.org/10.1177/07356331241240459

[3] Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educational Technology Research and Development*, *68*(4), 1903–1929. https://doi.org/10.1007/s11423-020-09793-2

[4] Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, Article 100017. https://doi.org/10.1016/j.caeai.2021.100017

[5] Valeva, P., Stoitsov, G., & Dimitrov, I. (2024). Artificial intelligence in advanced learning: Opportunities and challenges. *Mathematics and Informatics*, *67*(5), 561–579. https://doi.org/10.53656/math2024-5-7-art

[6] Futschek, G. (2006). Algorithmic thinking: The key for understanding computer science. In R. T. Mittermeir (Ed.), *Informatics education – The bridge between using and understanding computers* (pp. 159–168). Springer. https://doi.org/10.1007/11915355_15

[7] Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

[8] Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, *42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

[9] Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi report)*. California Academic Press.

[10] Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American*

Psychologist, 53(4), 449–455. https://doi.org/10.1037/0003-066X.53.4.449

[11] Brusilovsky, P. (2012). Adaptive hypermedia for education and training. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 46–68). Cambridge University Press. https://doi.org/10.1017/CBO9781139049580.006

[12] Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy*. Cambridge Adult Education.

[13] Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

[14] Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

[15] Tricco, A. C., Lillie, E., Zarin, W., O'Brien, et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

[16] Area9 Lyceum. (n.d.-c). *Why does Area9 Rhapsode measure learner confidence?* Retrieved October 14, 2026, from https://help.area9lyceum.com/biological-adaptive-model/why-does-area9-rhapsode-measure-learner-confidence/

[17] Area9 Lyceum. (n.d.-a). *Does Area9 Rhapsode support xAPI?* Retrieved October 14, 2026, from https://help.area9lyceum.com/integration/does-area9-rhapsode-support-xapi/

[18] Area9 Lyceum. (n.d.-b). *How is Area9 Rhapsode licensed?* Retrieved October 14, 2026, from https://help.area9lyceum.com/faqs/how-is-area9-rhapsode-licensed/

[19] Honke, N., & Becker-Genschow, S. (2025). Adaptive learning in bionics: Transforming science education. *Frontiers in Education, 10*, Article 1427083. https://doi.org/10.3389/feduc.2025.1427083

[20] Cornerstone. (n.d.). *Learning management system*. Retrieved October 14, 2026, from https://www.cornerstoneondemand.com/platform/learning-management-lms/

[21] Coursera. (n.d.). *Online business learning platform*. Retrieved October 14, 2026, from https://www.coursera.org/business

[22] Docebo. (n.d.). *AI-first LMS for enterprise learning and skills*. Retrieved October 14, 2026, from https://www.docebo.com/

[23] Litmos. (n.d.). *Litmos AI*. Retrieved October 14, 2026, from https://www.litmos.com/platform/litmos-ai/