ISSN 1981-416X
https://doi.org/10.7213/1981-416X.26.090.DS03EN

CC BY

ISSN

Dossier

Revista
Diálogo
Educacional

PUCPRESS

periodicos.pucpr.br/dialogoeducacional

Metacognition, Feedback, and Artificial Intelligence: A Didactic-Pedagogical Matrix for Learning Recalibration

Metacognição, feedback e inteligência artificial: uma matriz didático-pedagógica para a recalibração da aprendizagem

Rafael Castro de Souza [a]
Natal, RN, Brasil
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN)

Diogo Pereira Bezerra [b]
Ipanguaçu, RN, Brasil
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN)

Carla Katarina de Monteiro Marques [c]
Natal, RN, Brasil
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN)

How to cite: de SOUZA, R.; BEZERRA, D. P.; MARQUES, C. K. de M. Metacognition, Feedback, and Artificial Intelligence: A Didactic-Pedagogical Matrix for Learning Recalibration. *Revista Diálogo Educacional*, Curitiba, PUCPRESS, v. 26, n. 90, p. 1397-1421, jul./set. 2026. https://doi.org/10.7213/1981-416X.26.090.DS03EN

Abstract

This article proposes a didactic-metacognitive matrix to guide the use of educational technologies and artificial intelligence in transforming evidence of discrepancies between self-perception and performance into pedagogical feedback, student reflection, and learning recalibration. The study is theoretical-propositional, applied in nature, and exploratory-descriptive in purpose, articulating metacognition, diagnostic assessment, self-regulated learning, feedback, and teacher mediation. As a

[a] Doutorando em Ensino pelo Programa de Pós-Graduação em Ensino da Rede Nordeste de Ensino (RENOEN), e-mail: rafael.castro@ifrn.edu.br
[b] Doutor em Engenharia Química pela Universidade Federal do Ceará, e-mail: diogo.bezerra@ifrn.edu.br
[c] Doutor do em Engenharia de Teleinformática pela Universidade Federal do Ceará (2010) e pós-doutorado na UdeM (Department of Computer Engineering pela University Polytechnique of Montreal), e-mail: carla.marques@ifrn.edu.br

proof of concept, it uses a functional prototype of the Metacognitive Calibration Assisted Model (MCAM), which records, item by item, prospective judgment in the pre-test, objective performance in the test, and retrospective confidence in the post-test. The technical-computational demonstration, conducted with simulated profiles in a controlled scenario, verifies the feasibility of implementation and the logical-operational coherence of the MCAM, including recoding, independent normalization, signed discrepancy, fuzzy inference, and analytical visualizations. The matrix is organized into four movements: diagnose, provide feedback, reflect, and recalibrate. The demonstration results show that the computational chain is executable and interpretable, while pedagogical decision-making remains under teacher responsibility. It is concluded that the MCAM can support the organization of metacognitive evidence for formative practices when articulated with pedagogical mediation and the sociotechnical conditions of use.

Keywords: Metacognition. Pedagogical feedback. Artificial intelligence. Diagnostic assessment. Self-regulated learning.

Resumo

Este artigo propõe uma matriz didático-metacognitiva para orientar o uso de tecnologias educacionais e inteligência artificial na transformação de evidências de discrepância entre autopercepção e desempenho em feedback pedagógico, reflexão discente e recalibração da aprendizagem. O estudo é teórico-propositivo, de natureza aplicada e finalidade exploratório-descritiva, articulando metacognição, avaliação diagnóstica, aprendizagem autorregulada, feedback e mediação docente. Como prova de conceito, utiliza-se um protótipo funcional do Modelo Assistido de Calibração Metacognitiva (Metacognitive Calibration Assisted Model — MCAM), que registra, item a item, julgamento prospectivo no pré-teste, desempenho objetivo no teste e confiança retrospectiva no pós-teste. A demonstração técnico-computacional, realizada com perfis simulados em cenário controlado, verifica a viabilidade de implementação e a coerência lógico-operacional do MCAM, incluindo recodificação, normalização independente, discrepância assinada, inferência fuzzy e visualizações analíticas. A matriz organiza-se em quatro movimentos: diagnosticar, oferecer feedback, refletir e recalibrar. Os resultados da demonstração mostram que a cadeia computacional é executável e interpretável, enquanto a decisão pedagógica permanece sob responsabilidade docente. Conclui-se que o MCAM pode apoiar a organização de evidências metacognitivas para práticas formativas, desde que articulado à mediação pedagógica e às condições sociotécnicas de uso.

***Palavras-chave**: Metacognição. Feedback pedagógico. Inteligência artificial. Avaliação diagnóstica. Autorregulação da aprendizagem.*

1. Introduction

Learning assessment has historically been associated with the verification of student performance through grades, correct and incorrect answers, attendance, participation, and other observable indicators. Although these elements are relevant to educational monitoring, they do not exhaust the complexity of learning. Two students may obtain similar results in an assessment activity and yet differ in how they interpret their own performance: one may recognize their difficulties, while another may overestimate their understanding; one may demonstrate mastery but low confidence, while another may show low performance accompanied by high self-confidence.

This difference between demonstrated performance and students' self-perception leads directly to the field of metacognition. For Flavell (1979), metacognition involves the knowledge individuals have about their own cognitive processes, their metacognitive experiences, their goals, and the strategies mobilized when facing tasks. From this perspective, learning means not only acquiring content knowledge but also developing the capacity to monitor, evaluate, and regulate one's own learning process. Metcalfe (2009) emphasizes that metacognitive judgments can participate in the control of study decisions, while Persky, Lee, and Schlesselman (2020), by comparing perceived learning with objective knowledge gains, show that these measures are not equivalent. Self-perception therefore constitutes relevant evidence of the student's judgment, but it should not replace a measure of demonstrated performance.

The relationship between what students believe they know and what they actually demonstrate can be examined as a problem of metacognitive calibration. Greater calibration corresponds to closer alignment between judgment and performance evidence; divergence between these measures may take different directions and intensities, including overestimation and underestimation. The Dunning–Kruger effect is a specific manifestation of this problem, associated, in certain contexts, with greater overestimation among lower-performing participants (Kruger; Dunning, 1999; Pennycook et al., 2017). Accordingly, in this article the Dunning–Kruger effect is not used as a synonym for every metacognitive discrepancy or as a fixed diagnosis of the student. Mahmood's (2016) review further reinforces, in another domain, that subjective judgments and objective performance may diverge substantially.

In educational contexts, the accuracy of these judgments is pedagogically relevant because it can influence decisions about study, review, and help seeking. Dunlosky and Rawson (2012) demonstrated that monitoring errors, particularly overconfidence, can impair learning decisions and retention. At the other end of the calibration spectrum, underestimation also deserves attention, because adequate performance accompanied by an excessively negative judgment may require mediation different from that needed when low performance is accompanied by high confidence. Diagnostic assessment therefore becomes more informative when it jointly considers what students demonstrate and how they judge what they know.

The literature on the Dunning–Kruger effect in teaching and learning indicates that discrepancies between self-perception and performance have often been measured by combining self-assessments, objective tests, and statistical procedures. However, gaps remain in the development of specific educational technologies capable of supporting not only the diagnosis of these discrepancies but also their transformation into feedback, reflection, and pedagogical intervention (Souza et al., 2025). Thus, the challenge is not limited to identifying metacognitive miscalibration; it is also necessary to understand how to transform it into a formative opportunity.

Educational technologies and artificial intelligence can contribute to this process when they are used to organize evidence, identify patterns, and support more individualized feedback. Their incorporation into learning assessment, however, should not be understood as a replacement for the teacher. On the contrary, technology-assisted models need to be conceived as mediation tools subordinated to didactic intentionality and to the teacher's pedagogical interpretation. This position is consistent with Libâneo (2017), who understands teaching as intentional work articulated through objectives, methods, content, and assessment, and with Freire (1996), who argues that teaching is not the transfer of knowledge but the creation of conditions for its production.

The discussion becomes even more relevant in light of the advance of Artificial Intelligence (AI) in educational contexts. Contemporary debate on metacognition and learning requires pedagogical practices that move beyond both the mechanical transmission of information and purely classificatory assessment. In this setting, AI can expand the possibilities for monitoring, analysis, and feedback, provided that it is articulated with teacher education, cognitive ergonomics, educational management, and pedagogical mediation. The problem, therefore, is not merely technical but didactic: how can evidence of discrepancies between self-perception and performance be transformed into opportunities for reflection, feedback, and self-regulation?

Accordingly, the general objective of this study is to propose a didactic-metacognitive matrix to guide the use of educational technologies and artificial intelligence in transforming evidence of discrepancies between perceived knowledge and demonstrated performance into pedagogical feedback, student reflection, and learning recalibration. As an illustrative element, the study considers the logical and operational structure of a functional prototype based on pre-test, test, and post-test stages, capable of organizing situated evidence of overestimation, underestimation, and closer correspondence between judgment and performance.

2. Metacognition, Self-Perception, and Learning Calibration

Metacognition occupies a strategic place in contemporary discussions of learning because it enables an understanding of the relationship among knowledge, awareness, and regulation. In Flavell (1979), the concept encompasses different dimensions: metacognitive knowledge, metacognitive experiences, goals or tasks, and the actions or strategies mobilized by the individual. This formulation shifts learning away from an interpretation centered exclusively on outcomes toward a broader understanding in which students must also interpret their own cognitive pathway.

Metacognitive knowledge involves beliefs and judgments that individuals construct about themselves, the tasks they face, and the strategies available to address them. A student may believe that they master a particular topic, that a task is simple, or that a study strategy is sufficient even when subsequent performance does not confirm that perception. Likewise, a student may consider themselves unable to perform a task despite possessing sufficient knowledge to do so. In this sense, self-perception is a relevant dimension of learning, but it cannot be taken in isolation as evidence of competence.

The judgments students make about their own knowledge are not the same as demonstrated knowledge. Metcalfe (2009) discusses these judgments as part of metacognitive monitoring that can guide the control of study. Persky, Lee, and Schlesselman (2020) observed low correspondence between perceived learning and objective knowledge gains, while Finn and Tauber (2015) show that confidence can be influenced by subjective cues that do not always signal actual knowledge. This distinction supports the need to compare subjective and objective measures rather than use one as a substitute for the other.

In the metamemory literature, judgments of learning (JOLs) are prospective judgments made during or after a study experience to estimate future performance, especially the probability of recalling information (Koriat, 1997; Metcalfe, 2009). The MCAM pre-test is conceptually close to this field because it is prospective and conducted item by item, but it is not treated as a classic JOL: its question focuses on perceived ability or mastery in relation to the task before the objective response. The post-test, in turn, is retrospective and is closer to a confidence judgment about a response already given. This distinction avoids an inappropriate equivalence between metacognitive measures that differ in timing and object of judgment.

Within this comparison, calibration does not mean the complete absence of error, but rather the degree of correspondence between the judgment issued and the performance evidence in the context of a task. The discrepancy may be positive, negative, or close to zero and may vary in magnitude. This characterization allows overestimation,

underestimation, and closer correspondence to be treated as situated relational states that depend on the content, task, instrument, and assessment moment, rather than as permanent traits of the student.

Studies of the Dunning–Kruger effect constitute one specific strand of this discussion. Kruger and Dunning (1999) observed that lower-performing participants could have greater difficulty recognizing the inadequacy of their own responses. Pennycook et al. (2017) likewise found patterns of overestimation associated with performance in reasoning tasks, while their results also show variation in judgments across performance groups. For this reason, metacognitive calibration is adopted here as the broader phenomenon. The MCAM does not presuppose the occurrence of the Dunning–Kruger effect; rather, it represents relationships of greater or lesser correspondence, as well as the direction and intensity of discrepancy.

From this perspective, metacognitive discrepancy should not be reduced to a marker of individual deficiency. It should be understood as pedagogical evidence situated within a relationship among task, student, context, mediation, and feedback. A student may overestimate performance in one content area and be appropriately calibrated in another; may underestimate performance in higher-pressure assessment situations; or may modify self-perception after receiving qualified feedback. Calibration should therefore be understood as a dynamic process rather than as a fixed characteristic of the individual.

3. Metacognitive Feedback and Teacher Mediation

Diagnostic assessment makes it possible to gather information about prior knowledge, difficulties, gaps, and intervention needs. However, when limited to measuring performance, it may indicate that a student performed poorly without revealing whether the student recognizes that difficulty. This point is decisive for pedagogical practice. A student who answers incorrectly and knows that they did so requires one type of intervention; a student who answers incorrectly while believing they fully master the content requires another type of mediation.

Esomonu and Eleje (2020) discuss diagnostic assessment as a procedure capable of identifying difficulties and guiding pedagogical actions. From a metacognitive perspective, however, diagnosis must also incorporate students' self-perception. This makes it possible to observe the relationship between the result obtained and the judgment students make about their own learning. Assessment thus ceases to be merely a verification of performance and begins to produce evidence about how students interpret their formative pathway.

Metacognitive feedback emerges as an essential stage in this relationship. Identifying discrepancies between self-perception and performance is not enough. If the diagnosis is not transformed into comprehensible and pedagogically oriented feedback, it risks remaining merely technical data. Metacognitive feedback should make explicit to students the distance between what they believed they knew, what they demonstrated that they knew, and what they need to reorganize in their study strategies.

In this sense, feedback should not operate solely as error correction. It should function as mediation so that students can understand their own judgment criteria, recognize gaps, revise strategies, and develop greater awareness of their learning process. This understanding is consistent with Libâneo (2017), who places assessment within intentional pedagogical practice, and with Luckesi (2011, pp. 13–15), who understands assessment as a source of information that supports decisions and the redirection of pedagogical action.

The relationship among monitoring, feedback, and revision of judgments is also connected to self-regulated learning. Panadero's (2017) synthesis shows that models of self-regulation integrate processes of monitoring, control, assessment, and reflection in recursive cycles. From this perspective, feedback can provide external criteria that allow students to compare their judgments with evidence, but it does not, by itself, guarantee recalibration. Revising a judgment depends on the student's own activity, the characteristics of the task, and the conditions of pedagogical mediation. In this scenario, the contribution of educational technologies and artificial intelligence is to support the organization of evidence and the identification of patterns that can inform teacher action. The pedagogical meaning of

feedback, however, remains dependent on human mediation. A system can indicate that there is a discrepancy between self-perception and performance; it is up to the teacher to interpret that information, contextualize the situation, consider the student's trajectory, and decide which interventions are most appropriate.

4. Educational Technologies, Artificial Intelligence, and Sociotechnical Conditions

Digital technologies, learning analytics, and AI broaden the possibilities for producing, organizing, and interpreting educational evidence, but their value depends on articulation with explicit pedagogical models. Banihashem et al. (2025), in a systematic review of 93 studies, show that learning analytics results become more useful for formative assessment when connected to processes that clarify where the student is, where the student needs to go, and how progress can be made. In parallel, the meta-systematic review by Bond et al. (2024) indicates that research on AI in higher education requires simultaneous attention to methodological rigor, ethics, and the context of application.

Teacher education is a central condition for this integration. Børte et al. (2023), in synthesizing studies on the use of technology in formative assessment, identify three prerequisites: a clear definition of formative purpose, alignment between the digital tool and pedagogical practice, and data literacy to interpret information and convert it into actions that improve learning. For the MCAM, this evidence reinforces that calibration indicators acquire educational meaning only when teachers understand what they represent and decide how to use them.

The cognitive dimension must also be considered. Skulmowski and Xu (2022) show that digital environments can introduce extraneous cognitive load, but they emphasize that the effects of technological elements should be analyzed in relation to learning objectives and the pedagogical benefits they can produce. In the MCAM, this implies presenting results in a legible and information-efficient manner, preventing the quantity of indicators, abbreviations, or visualizations from becoming an additional burden for students and teachers.

Institutional adoption of analytical technologies also involves governance, processes, and different actors. Márquez et al. (2024), in a systematic review of learning analytics in higher education, identify factors distributed across institutional, pedagogical, technical, and ethical dimensions. The incorporation of the MCAM therefore should not be conceived merely as installing software, but as integrating evidence into monitoring routines, interpretation criteria, data protection practices, and pedagogical decisions.

Accordingly, a didactic-metacognitive matrix supported by computational resources needs to consider at least four sociotechnical conditions: teacher education and data literacy; cognitive ergonomics and clarity of representations; institutional organization of data and processes; and pedagogical mediation capable of transforming indicators into formative feedback. Peer-reviewed literature converges on an interpretation in which technology can expand diagnostic and monitoring capacity, but its contribution depends on how results are interpreted, contextualized, and incorporated into practice (Børte et al., 2023; Márquez et al., 2024; Bond et al., 2024; Banihashem et al., 2025).

5. Methodological Approach

The study is theoretical and propositional, applied in nature, and exploratory and descriptive in purpose, accompanied by a technical and computational demonstration of the Metacognitive Calibration Assisted Model (MCAM) using simulated profiles constructed for controlled testing. The computational stage was designed to test implementation feasibility and the logical and operational coherence among the instrument, coding, normalization, discrepancy calculation, fuzzy inference, persistence, and visualization. The prototype therefore functions as a proof of concept for the proposed analytical chain. Empirical evaluation of the model in real educational contexts is part of a subsequent stage of the research program and is currently underway; this article focuses on conceptual specification and technical and computational feasibility.

The prototype was implemented in Java, with a Swing interface, a TCP-based client-server architecture, and relational persistence. The teacher registers an assessment unit composed of three moments linked to the same content: (i) a prospective judgment of mastery/ability (pre-test); (ii) an objective performance question (test); and (iii) a retrospective confidence judgment (post-test). In the student interface, responses are collected item by item and returned to the server for storage and aggregation. This granularity makes it possible to compare pre-test × test and post-test × test separately, without assuming that the two metacognitive judgments are conceptually identical measures. The mathematical specification is presented in Appendix A, while the execution architecture and corresponding interfaces are presented in Appendix B.

The A–D alternatives used for prospective and retrospective judgments are recoded as 0, 1, 2, and 3 for analytical purposes. Let p_i be the code assigned to the judgment on item i, n the number of items, R the objective score obtained, and R_max the maximum possible test score. The measures are normalized independently to the [0,1] interval, preventing the comparison from depending on the raw scale chosen for the test. The same equations are applied separately to the pre-test and post-test.

$$P = \Sigma_i\, p_i, \quad \text{with } p_i \in \{0,1,2,3\}$$
$$P_N = P / (3n) \qquad R_N = R / R_max$$
$$d = P_N - R_N \qquad m = |d|$$

The sign of d expresses direction: $d > 0$ indicates overestimation, $d < 0$ indicates underestimation, and values close to zero indicate closer correspondence between judgment and performance. The mathematical magnitude is $m = |d|$. At this stage, the article treats discrepancy as a continuous variable and does not establish a universal range of "acceptable calibration"; any tolerance applied in an educational context should be examined according to the instrument, task, and context. The complete specification of the variables, including P_N,PRE, P_N,POST, d_PRE, and d_POST, is provided in Appendix A.

This distinction between direction and magnitude also appears in the literature on metacognitive accuracy: Persky, Lee, and Schlesselman (2020) treat bias as the difference between prediction and performance and absolute bias as the absolute value of that difference, distinguishing, respectively, the direction and magnitude of judgment error.

In addition to signed discrepancy, the MCAM uses fuzzy inference to represent magnitude in linguistic terms. The normalized inputs are KnowledgeReal = R_N and SelfPerception = P_N; the output, Discrepancy, uses the terms Low, Moderate, and High. The rule base contains nine rules that assign low discrepancy to aligned combinations (Low/Low, Medium/Medium, and High/High), moderate discrepancy to adjacent combinations, and high discrepancy to opposite extremes. The inference system uses AND = MIN, ACT = MIN, and ACCU = MAX, with Center of Gravity (COG) defuzzification and a default value of zero. Because the rule matrix is symmetric, the fuzzy output represents magnitude; direction continues to be provided by the sign of d. The vertices of the membership functions and the equations of the inference mechanism are presented in Appendix A; the structure of the nine-rule base is described above.

Prototype outputs include normalized scatter visualization, membership functions, and the defuzzified result. The software also maintains an auxiliary ROC analysis module to explore the sensitivity of thresholds defined by the teacher: for each cutoff c between 0.01 and C, the system calculates binary predictions, a confusion matrix, TPR, and FPR. In the demonstration, its function is to examine how changes in the threshold modify true- and false-positive rates when compared with an externally assigned reference class; the choice of threshold remains under teacher control, and the system allows the sensitivity of the results to be compared across the simulated cutoffs. The normalized scatter plot is presented in Figure 1, while Appendix B provides examples of the fuzzy inference output and the auxiliary ROC module.

The main demonstration uses ten synthetic profiles deterministically defined in a ten-item scenario. The recoded judgment total ranges from 0 to 30 points, and objective performance in the demonstration ranges from 0 to 40 points. The profiles were intentionally constructed to cover patterns of closer correspondence, overestimation, and underestimation at different magnitudes and thereby subject the computational chain to contrasting, controllable, and traceable cases. The mathematical and fuzzy parameters used in the computational chain are presented in Appendix A.

6. Results and Discussion: A Didactic-Metacognitive Matrix for Learning Recalibration

As a theoretical and propositional result, a didactic-metacognitive matrix organized into four movements is presented: diagnose, provide feedback, reflect, and recalibrate. The matrix is based on the assumption that learning assessment should consider not only objective performance but also how students perceive, interpret, and regulate their own knowledge. The prototype provides a computational demonstration of how part of this evidence can be organized; transforming the results into pedagogical action remains dependent on teacher mediation.

6.1 Diagnose: Identifying the Relationship between Self-Perception and Performance

The first movement of the matrix consists of diagnosing the relationship between judgment and performance. At this stage, the objective is not only to identify whether a student answered correctly or incorrectly, but also to examine whether the judgment made about their own knowledge approaches or diverges from the available objective evidence. The diagnosis is therefore relational and situated: it describes a combination observed in a particular task and moment, not a permanent characteristic of the student.

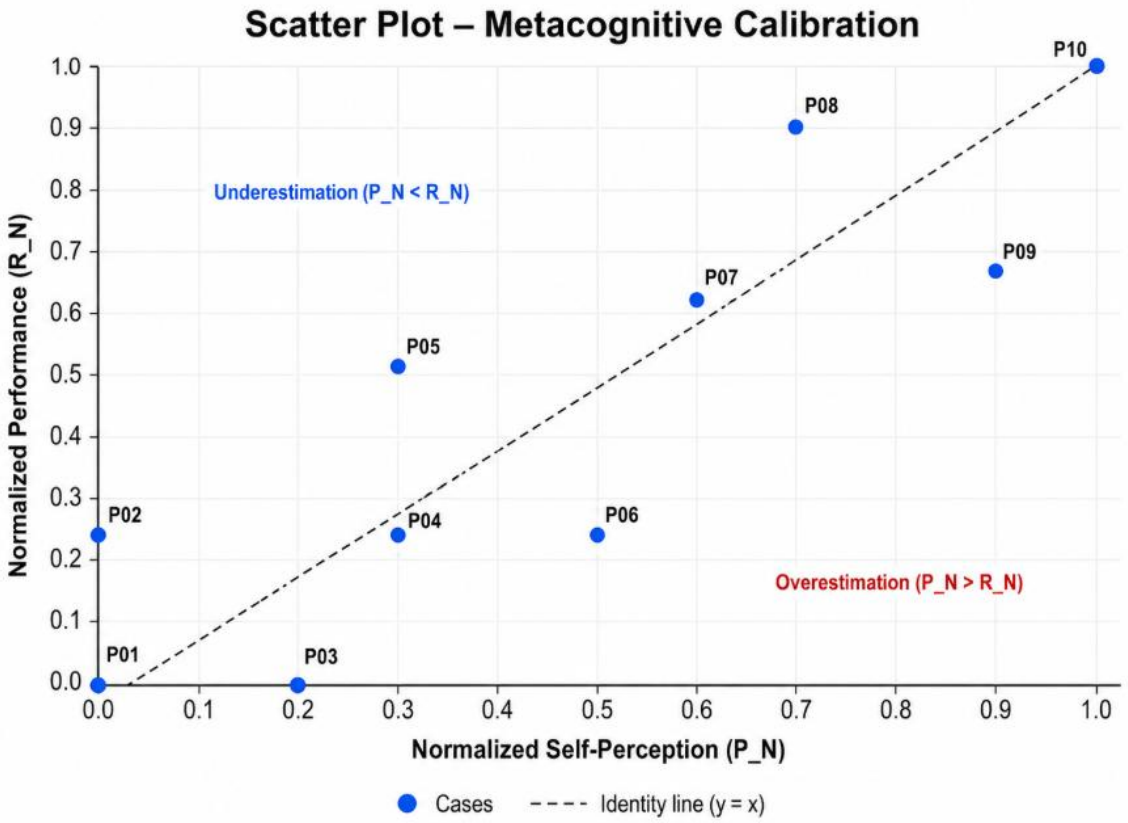
In the prototype, the three moments provide distinct and complementary measures. The pre-test records a prospective judgment of ability or mastery before the question; the test records objective performance on the same item; and the post-test records a retrospective confidence judgment after the response. Pre-test × test and post-test × test comparisons can therefore be examined separately, avoiding treatment of the two judgments as if they were the same repeated measure.

The relationship between the measures may take three general directions. Overestimation occurs when $P_N > R_N$, underestimation when $P_N < R_N$, and closer correspondence when the measures approach one another. These categories describe the observed relationship and do not constitute fixed labels. The intensity of the divergence also matters: two cases of overestimation may have very different magnitudes and therefore require different pedagogical interpretations.

Figure 1 presents the normalized visualization of the ten simulated profiles described in the methodology. The horizontal axis represents the normalized prospective judgment (P_N), and the vertical axis represents normalized objective performance (R_N). The identity line $y = x$ represents exact correspondence between the measures. Because each point represents a simulated case, points below the line indicate overestimation ($P_N > R_N$), points above the line indicate underestimation ($P_N < R_N$), and points on or near the line indicate closer correspondence.

Distance from the identity line can be interpreted together with the signed discrepancy $d = P_N - R_N$. The sign indicates direction and $|d|$ indicates mathematical magnitude; fuzzy inference adds a linguistic representation of magnitude through transparent rules and explicit parameters. The demonstration does not establish a rigid calibration range or transform point position into a definitive diagnosis. Appendix A allows the calculation to be reconstructed from item-by-item responses to the defuzzified output, and Appendix B shows examples of execution of the fuzzy module and the auxiliary ROC module.

Figure 1 – Normalized scatter plot produced by the prototype for ten simulated profiles.



Source: Authors' own work.

6.2 Feedback: Transforming Diagnosis into Pedagogical Feedback

The second movement consists of transforming diagnosis into metacognitive feedback. Identifying discrepancy acquires educational meaning only when it is communicated to the student in a comprehensible, contextualized, and learning-oriented manner. Metacognitive feedback is not limited to stating whether an answer is correct or incorrect; it makes explicit the relationship between the student's judgment and their demonstrated performance.

In cases of overestimation, feedback can help students recognize that their confidence did not correspond to the result obtained. This should not be done punitively, but as an opportunity to review self-assessment criteria, study strategies, and understanding of the content. In cases of underestimation, feedback can support recognition of successful strategies and strengthen academic confidence. When there is closer correspondence between judgment and performance, feedback can reinforce monitoring practices that contributed to a better fit between perception and performance.

In the implementation demonstrated here, the computational core organizes measures, performs fuzzy inference, and produces visualizations that support the teacher's formulation of feedback. Within the matrix, artificial intelligence resources can expand this support by providing initial feedback suggestions and organizing evidence; however, the final formulation of the feedback must remain the responsibility of the teacher, taking into account the class context, the nature of the content, the student's circumstances, and the formative objectives of the assessment.

6.3 Reflect: Transforming Feedback into Metacognitive Activity

The third movement of the matrix is reflection. Feedback becomes metacognitive only when it prompts students to think about their own judgment criteria and learning strategies. Without this stage, technology risks merely presenting indicators without promoting changes in students' awareness of their learning process.

Reflection can be guided by questions such as: "What did I believe I knew before the activity?", "What did my performance show?", "Where did my perception approach or diverge from the result?", "What strategy did I use?", "What strategy do I need to modify?", and "What evidence indicates that I did or did not understand the content?". Such questions encourage students to reconstruct the relationship among perception, performance, and action.

Within the matrix, artificial intelligence resources can support the formulation of reflective questions, the organization of study pathways, and the synthesis of monitoring records based on evidence produced by the analytical core. This layer is conceived as support for reflective activity while preserving the active participation of the student and pedagogical mediation. The prototype presented here provides the analytical core that organizes the measures, calculates discrepancy, performs fuzzy inference, and produces visualizations to support this movement.

6.4 Recalibrate: Progressively Adjusting Perception and Performance

The fourth movement is learning recalibration. Recalibration means fostering, over successive cycles, a progressively better-informed relationship between what students judge that they know and what they are able to demonstrate. It is a pedagogical purpose of the matrix, not an automatic effect produced by the software or by a single instance of feedback.

Comparison among the pre-test, test, and post-test makes it possible to observe, within the same assessment cycle, differences among prospective judgment, objective performance, and retrospective confidence. When this structure is repeated over time, it provides a basis for tracking changes in judgments and performance. Subsequent empirical studies can therefore examine the relationship among feedback, reflection, and recalibration longitudinally, using indicators produced by the MCAM chain itself.

Recalibration is connected to self-regulation because metacognitive judgments can participate in decisions about study, review, help seeking, and changes in strategy (Metcalfe, 2009; Panadero, 2017). In this sense, the matrix does not assume that perception and performance will spontaneously converge; rather, it organizes conditions in which discrepancies can become objects of pedagogical analysis and can inform new cycles of monitoring and decision-making.

6.5 Sociotechnical Conditions and Synthesis of the Proposed Matrix

Applying the proposed matrix requires sociotechnical conditions that extend beyond the existence of the prototype. The first is teacher education associated with data literacy. Børte et al. (2023) show that the use of technologies in formative assessment depends on teachers who are able to interpret the information produced and translate it into learning-oriented decisions. In the MCAM, this means understanding the pedagogical meaning of P_N, R_N, d, and the fuzzy outputs, rather than merely operating the interface.

The second condition concerns cognitive ergonomics. Skulmowski and Xu (2022) propose that the design of digital environments balance cognitive load with the intended learning outcomes. For this reason, MCAM outputs should prioritize legibility, visual hierarchy, and actionable information: mathematical formalization can remain available for audit, while everyday pedagogical use should provide representations simple enough to support interpretation.

The third condition concerns transforming data into formative evidence. Banihashem et al. (2025) show that learning analytics gains pedagogical value when its results are aligned with formative assessment processes and help

clarify the student's current situation, learning objectives, and pathways for progress. This logic is compatible with the MCAM insofar as discrepancy is positioned as evidence for diagnosis and feedback rather than as a final classification.

The fourth condition is institutional organization for adoption and governance. Márquez et al. (2024) identify learning analytics adoption factors distributed across actors, pedagogical and institutional processes, technical limitations, and ethical considerations. Sustaining the MCAM therefore requires monitoring routines, data-access criteria, clearly defined responsibilities, and spaces in which results can be discussed pedagogically.

These conditions reinforce that artificial intelligence should not be conceived as an isolated solution. The meta-review by Bond et al. (2024) highlights the need for greater attention to ethics, context, collaboration, and methodological rigor in educational research involving AI. In the MCAM, this is translated into a support architecture: the analytical component organizes evidence and expands the capacity for interpretation, while interpretation and pedagogical decision-making remain situated in the relationship among teacher, student, and context. The proposed didactic-metacognitive matrix can be summarized in four articulated movements, as shown in Table 1.

Table 1 – Didactic-metacognitive matrix for feedback and recalibration

Movement	Pedagogical Purpose	Evidence Used	Possible Technological Support	Teacher's Role
Diagnose	Identify the relationship between self-perception and performance	Pre-test, test, post-test, self-assessment, and performance	Data organization, visualizations, and pattern identification	Interpret the student profile and contextualize the results
Provide Feedback	Transform diagnosis into formative feedback.	Discrepancy between perception and performance	Feedback suggestions, reports, and alerts	Produce contextualized, ethical, and pedagogical feedback
Reflect	Stimulate metacognitive awareness	Reflective responses, justifications, and reported strategies	Guiding questions and study pathways	Mediate the student's critical analysis of their own process
Recalibrate	Foster progressive adjustment between judgment and performance across cycles	Longitudinal comparison across cycles, performance, and new judgments	Longitudinal monitoring and comparison across moments	Plan interventions and monitor changes in metacognitive judgment

Source: Authors' own work.

The matrix seeks to shift assessment from a predominantly classificatory logic toward a formative logic. Its contribution does not lie in automating teacher judgment, but in organizing evidence that can enhance pedagogical mediation. Students cease to be assessed only by their outcomes and are also monitored in relation to how they interpret their own learning. In this sense, metacognitive diagnosis becomes the starting point for feedback, reflection, and recalibration.

7. Final Considerations

This article proposed a didactic-metacognitive matrix to guide the use of educational technologies and artificial intelligence in transforming evidence of discrepancies between perceived knowledge and demonstrated performance into pedagogical feedback, student reflection, and learning recalibration. The proposal is based on the understanding that educational assessment should not be limited to measuring objective performance, because the way students judge and interpret their own knowledge adds relevant information to diagnostic interpretation.

The matrix is organized into four movements: diagnose, provide feedback, reflect, and recalibrate. To diagnose means to examine the relationship between judgment and performance; to provide feedback means to transform this

evidence into pedagogical feedback; to reflect means to foster analysis of the criteria and strategies used; and to recalibrate means to seek, across new cycles, judgments that are progressively better informed by experience and the available evidence.

As a contribution, the study shifts the use of technology from a merely classificatory function toward a mediating pedagogical function. Technology-assisted models can support teachers in organizing evidence, visualizing patterns, and identifying metacognitive calibration profiles. However, such models do not replace teacher judgment and should not be used to label students. Their purpose should be to support pedagogical decisions, formative feedback, and individualized monitoring strategies.

The study also shows that the educational use of AI requires sociotechnical conditions. Teacher education, cognitive ergonomics, data organization, and institutional management are fundamental dimensions if technology is to contribute effectively to learning. Without these conditions, assessment risks being reduced to automated indicators. With them, it becomes possible to develop assessment practices that are more diagnostic, reflective, and oriented toward self-regulation.

It is concluded that artificial intelligence resources can contribute to learning when they are subordinated to pedagogical intentionality and teacher mediation. In the prototype demonstrated here, fuzzy inference is the computational artificial intelligence component used to integrate normalized measures into an interpretable representation of discrepancy magnitude. This information is conceived as support for teacher mediation, which is responsible for transforming evidence into feedback and guiding cycles of reflection and recalibration. The contribution of the MCAM lies in making the relationship between judgment and performance examinable and making it available in a traceable form to support formative decisions and self-regulatory processes.

References

BANIHASHEM, Seyyed Kazem *et al. Optimizing formative assessment with learning analytics. Review of Educational Research, OnlineFirst*, 2025.

BOND, Melissa *et al. A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education*, v. 21, art. 4, 2024.

BØRTE, Kristin *et al. Prerequisites for teachers' technology use in formative assessment practices: a systematic review. Educational Research Review*, v. 41, art. 100568, 2023.

DUNLOSKY, John; RAWSON, Katherine *A. Overconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. Learning and Instruction*, v. 22, n. 4, p. 271-280, 2012.

ESOMONU, N. P. M.; ELEJE, L. I. *Effect of diagnostic testing on students' achievement in secondary school quantitative economics. World Journal of Education*, v. 10, n. 3, p. 178-187, 2020.

FINN, Bridgid; TAUBER, Sarah K. *When confidence is not a signal of knowing: how students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. Educational Psychology Review*, v. 27, n. 4, p. 567-586, 2015.

FLAVELL, John H. *Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. American Psychologist*, v. 34, n. 10, p. 906-911, 1979.

FREIRE, Paulo. Pedagogia da autonomia: saberes necessários à prática educativa. São Paulo: Paz e Terra, 1996.

KORIAT, Asher. *Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. Journal of Experimental Psychology: General*, v. 126, n. 4, p. 349-370, 1997.

KRUGER, Justin; DUNNING, David. *Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology*, v. 77, n. 6, p. 1121-1134, 1999.

LIBÂNEO, José Carlos. Didática. São Paulo: Cortez, 2017.

LUCKESI, Cipriano Carlos. Avaliação da aprendizagem: componente do ato pedagógico. São Paulo: Cortez, 2011.

MAHMOOD, Khalid. *Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. Communications in Information Literacy*, v. 10, n. 2, p. 199-213, 2016.

MÁRQUEZ, Lucía *et al. Adoption of learning analytics in higher education institutions: a systematic literature review. British Journal of Educational Technology*, v. 55, n. 2, p. 439-459, 2024.

METCALFE, Janet. *Metacognitive judgments and control of study. Current Directions in Psychological Science*, v. 18, n. 3, p. 159-163, 2009.

PANADERO, Ernesto. *A review of self-regulated learning: six models and four directions for research. Frontiers in Psychology*, v. 8, art. 422, 2017.

PENNYCOOK, Gordon; ROSS, Robert M.; KOEHLER, Derek J.; FUGELSANG, Jonathan A. *Dunning-Kruger effects in reasoning: theoretical implications of the failure to recognize incompetence. Psychonomic Bulletin & Review*, v. 24, p. 1774-1784, 2017.

PERSKY, Adam M.; LEE, Edward; SCHLESSELMAN, Lauren S. *Perception of learning versus performance as outcome measures of educational research. American Journal of Pharmaceutical Education*, v. 84, n. 7, art. 7782, 2020.

SKULMOWSKI, Alexander; XU, Kate Man. *Understanding cognitive load in digital and online learning: a new perspective on extraneous cognitive load. Educational Psychology Review*, v. 34, p. 171-196, 2022.

SOUZA, Rafael Castro de *et al.* Efeito Dunning-Kruger no processo de ensino e aprendizagem: uma revisão sistemática da literatura. Boletim de Conjuntura (BOCA), Boa Vista, v. 22, n. 64, 2025.

Editor Responsável/Editor-in-Chief: Alboni Marisa Dudeque Pianovski Vieira

Recebido/Received: 10.05.2026 / 05.10.2026

Aprovado/Approved: 15.08.2026 / 08.15.2026

APPENDIX A – COMPUTATIONAL SPECIFICATION OF THE MCAM

A.1 Abbreviations and Symbols: How to Read the Equations

Table A1 summarizes the notation used in MCAM processing and distinguishes the normalized measures, signed discrepancy, its mathematical magnitude, and the fuzzy output. This identification also seeks to prevent quantities with different functions—such as signed discrepancy d, its mathematical magnitude |d|, and fuzzy output F—from being interpreted as equivalent.

Table A1 – Abbreviations, symbols, and their respective functions in MCAM processing.

Abbreviation/symbol	Meaning	Interpretation in the MCAM
MCAM	Metacognitive Calibration Assisted Model	Computational model supporting metacognitive calibration.
Pre-test / Test / Post-test	prospective / objective performance / retrospective	Pre-test: judgment before the item; Test: objective performance; Post-test: confidence after the item.
n; p_i	number of items; judgment code for item i	p_i: A=0, B=1, C=2, D=3; n: number of items.
P_PRE; P_POST; P	sums of judgments	Sum of the codes in the pre-test or post-test.
P_N	normalized perception	P/(3n), therefore ranging from 0 to 1.
R; R_max; R_N	score obtained; objective maximum; normalized performance	R_N=R/R_max, also within the 0–1 interval.
d; |d|	signed discrepancy; mathematical magnitude	d=P_N−R_N: sign=direction; |d|=distance without direction.
μ; μ_max	fuzzy membership; highest membership	Values from 0 to 1 expressing membership in the linguistic sets.
α_j	strength of rule j	Lowest membership degree among the conditions that activate a rule.

Abbreviation/symbol	Meaning	Interpretation in the MCAM
F	defuzzified fuzzy output	Numerical value obtained by COG; represents fuzzy magnitude, not direction.
MIN / MAX; COG	operators; Center of Gravity	MIN combines/activates; MAX aggregates; COG produces F.
TP / FP / TN / FN	confusion matrix	True/false positive and true/false negative at each cutoff.
TPR / FPR	true- and false-positive rates	The y and x coordinates, respectively, of each ROC point.
C; ε	maximum cutoff; numerical tolerance	C limits the cutoffs; $\varepsilon=10^{-9}$ helps mitigate floating-point comparison issues.
AI	Artificial intelligence	In the prototype, fuzzy inference is the AI component effectively used in the analysis.
JOL / JOLs	Judgment(s) of Learning	Prospective learning judgment in the metamemory literature; the pre-test is conceptually close to this field but is not treated as a classic JOL.
TCP / DAO	Transmission Control Protocol / Data Access Object	TCP provides client-server communication; DAO is the layer that persists and retrieves data.
ROC	Receiver Operating Characteristic	Sensitivity analysis of thresholds using TPR and FPR rates.

Source: Authors' own work.

A.2 From Raw Responses to a Common 0-to-1 Scale

Considering n as the total number of questions, the cumulative self-perception scores in the pre-test and post-test are obtained by summing the recoded values for each item:

$$P_{\mathrm{PRE}} = \sum_{i=1}^{n} p_{\mathrm{PRE},i} \qquad P_{\mathrm{POST}} = \sum_{i=1}^{n} p_{\mathrm{POST},i}$$

In these expressions, i denotes the question index, n corresponds to the total number of questions, and the pre-test and post-test terms represent, respectively, the recoded values of the student's responses for question i, adopting A = 0, B = 1, C = 2, and D = 3. The total objective test score is calculated analogously:

$$R = \sum_{i=1}^{n} r_i$$

where a correct response receives 4 points and an incorrect response receives 0 points. Because the self-perception and performance scales have different ranges, comparison is performed after independent normalization:

$$P_{N,\mathrm{PRE}} = \frac{P_{\mathrm{PRE}}}{3n} \qquad P_{N,\mathrm{POST}} = \frac{P_{\mathrm{POST}}}{3n} \qquad R_N = \frac{R}{R_{\max}}$$

Metacognitive discrepancy is then calculated as the difference between normalized self-perception and normalized objective performance:

$$d_{\mathrm{PRE}} = P_{N,\mathrm{PRE}} - R_N \qquad d_{\mathrm{POST}} = P_{N,\mathrm{POST}} - R_N$$

A.3 Discrepancy: Direction and Magnitude

After normalization, perception and performance lie within the same [0,1] interval. Discrepancy is calculated as the difference between normalized perception and normalized performance:

$$d = P_N - R_N$$

In this expression, d is a signed measure. When d > 0, perception is higher than demonstrated performance, characterizing overestimation. When d < 0, perception is lower than performance, characterizing underestimation. Values equal or close to zero represent closer correspondence between judgment and performance. Thus, the sign of d indicates the direction of divergence between the two measures, not merely its size. To represent only the distance between the two measures, regardless of direction, the absolute value is used:

$$m = |d|$$

The value m removes the sign and retains only the mathematical magnitude of discrepancy. Thus, two situations with d = +0.30 and d = −0.30 have the same mathematical distance (m = 0.30), although they have opposite directions: the former corresponds to overestimation and the latter to underestimation. This distinction is important because direction and magnitude provide complementary information for interpreting metacognitive calibration.

A.4 Fuzzy Inference of Discrepancy Magnitude

The fuzzy system receives the two previously normalized measures as inputs:

KnowledgeReal = R_N; SelfPerception = P_N

Each input is represented by the linguistic terms Low, Medium, and High. The output variable, named Discrepancy, uses the terms Low, Moderate, and High. For each value between 0 and 1, the membership functions establish the degree μ to which that value belongs to each fuzzy set. A single value may therefore simultaneously have different degrees of membership in neighboring sets, allowing gradual transitions to be represented instead of relying exclusively on rigid boundaries.

Fuzzy inference depends on the definition of the membership functions associated with the input and output variables. Table A2 presents the points that delimit these functions in the configuration used by the prototype.

Table A2 – Parameterization of the membership functions used in the MCAM fuzzy system.

Variable	Term	Membership-function points (x, μ)
KnowledgeReal	*Low*	*(0, 1); (0.1, 1); (0.2, 1); (0.3, 0)*
KnowledgeReal	*Medium*	*(0.2, 0); (0.3, 1); (0.4, 1); (0.5, 1); (0.6, 1); (0.7, 0)*
KnowledgeReal	*High*	*(0.6, 0); (0.7, 1); (0.8, 1); (0.9, 1); (1, 1)*
SelfPerception	*Low*	*(0, 1); (0.1, 1); (0.2, 1); (0.3, 0)*
SelfPerception	*Medium*	*(0.2, 0); (0.3, 1); (0.4, 1); (0.5, 0)*
SelfPerception	*High*	*(0.4, 0); (0.5, 1); (0.6, 1); (0.7, 1); (0.8, 1); (0.9, 1); (1, 1)*

Discrepancy	*Low*	*(0, 1); (0.1, 1); (0.2, 1); (0.3, 0)*
Discrepancy	*Moderate*	*(0.2, 0); (0.3, 1); (0.4, 1); (0.5, 0)*
Discrepancy	*High*	*(0.4, 0); (0.5, 1); (0.6, 1); (0.7, 1); (0.8, 1); (0.9, 1); (1, 1)*

Source: Authors' own work.

The points presented in Table A2 define how normalized values are distributed among the linguistic terms. For the inputs, performance and self-perception are represented by Low, Medium, and High; for the output, discrepancy is represented by Low, Moderate, and High. Partial overlap between the functions allows a single value to participate, to different degrees, in neighboring sets, preserving the gradual nature of fuzzy inference. The fuzzy output represents discrepancy magnitude, while its direction—as overestimation or underestimation—remains determined by the sign of the signed discrepancy. For each rule j, its activation strength is calculated as:

$$\alpha_j = \text{MIN}[\ \mu_Aj(R_N),\ \mu_Bj(P_N)\ ]$$

In this expression, μ_Aj(R_N) and μ_Bj(P_N) are the membership degrees of the two inputs in the sets specified by rule j. The MIN operator selects the smaller of these values. Thus, the extent to which a rule participates in the result is limited by the condition with the lower degree of membership. After the rules are activated, their contributions are combined into a single output function:

$$\mu_D(z) = \text{MAX_j}\ \{\ \text{MIN}[\ \alpha_j,\ \mu_Cj(z)\ ]\ \}$$

Here, μ_Cj(z) represents the membership function of the consequent of rule j. The MIN operator limits this function to the activation strength α_j, while the MAX operator combines the contributions of the rules into a single fuzzy distribution. At this stage, the result is not yet a single number but an aggregated discrepancy membership function. Finally, this distribution is converted into a numerical value using the Center of Gravity (COG) method:

$$F = [\int z \cdot \mu_D(z)\ dz] \,/\, [\int \mu_D(z)\ dz]$$

In this expression, F corresponds to the defuzzified value and μ_D(z) represents the membership degree of the aggregated output. The COG method determines the center of gravity of this distribution, producing a numerical value for fuzzy discrepancy. This value can then be interpreted together with the membership degrees of the Low, Moderate, and High terms without replacing the directional information provided by d. The configuration used by the prototype is:

$$\text{AND} = \text{MIN};\ \text{ACT} = \text{MIN};\ \text{ACCU} = \text{MAX};\ \text{METHOD} = \text{COG};\ \text{DEFAULT} = 0$$

These parameters make the inference sequence explicit: the minimum operator is used for conjunction and rule activation, the maximum operator for output aggregation, and the center of gravity for defuzzification. The default value of zero is used when no contribution is capable of producing a different output. This specification directly connects the mathematical formulation to the procedure executed by the system.

A.5 ROC Curve as Sensitivity Analysis of Cutoffs

The ROC module is used in the MCAM as an auxiliary resource to examine classification sensitivity to different cutoff values. The teacher establishes a maximum threshold C, and the program successively evaluates cutoffs of 0.01, 0.02, 0.03, and so forth, up to C. At each cutoff, the signed discrepancy is converted into a binary overestimation decision, where c represents the cutoff being examined:

$$d + \varepsilon \geq c; \varepsilon = 1 \times 10^{-9}$$

The term ε is a small computational tolerance used to reduce comparison problems arising from floating-point number representation. Therefore, ε is not a pedagogical or metacognitive threshold; its function is exclusively numerical. The classification produced at each cutoff is compared with the reference class used in the controlled exercise. This comparison yields true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values. From these, the true-positive rate and false-positive rate are calculated:

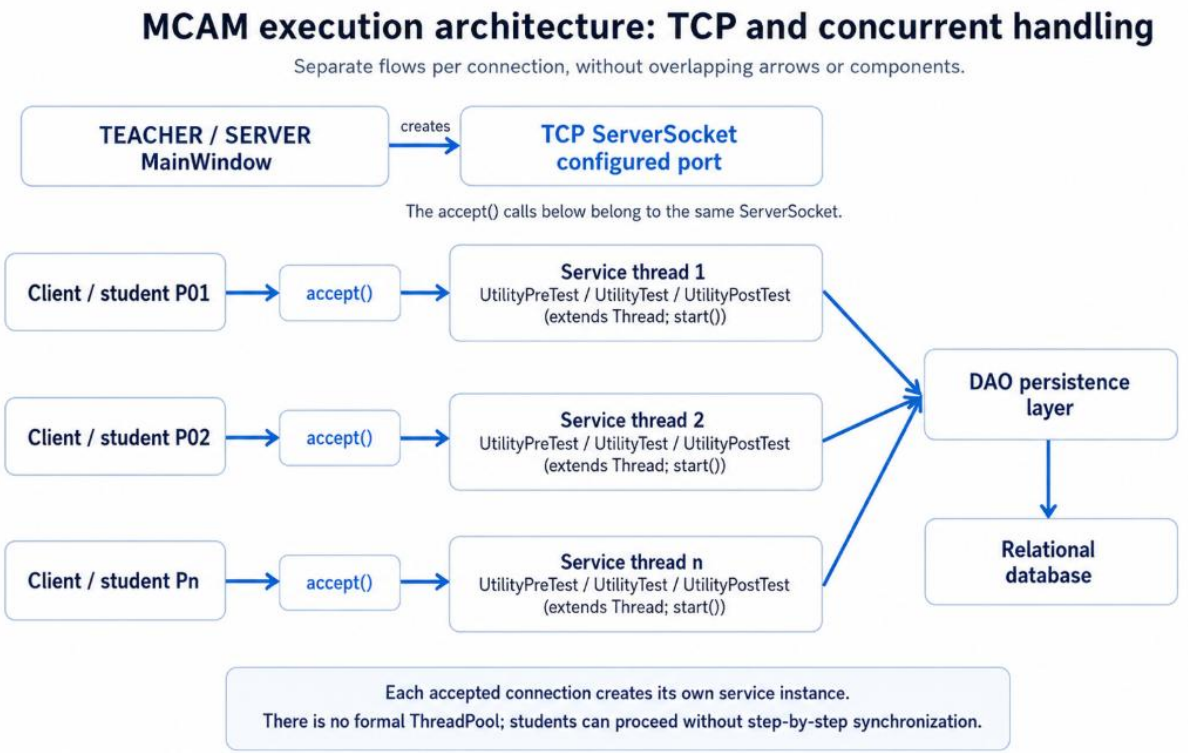
$$TPR = TP / (TP + FN); FPR = FP / (FP + TN)$$

TPR represents the proportion of reference positives correctly identified, while FPR represents the proportion of reference negatives classified as positive.

APPENDIX B – PROTOTYPE ARCHITECTURE, INTERFACES, AND VISUALIZATIONS

Figure B1 presents the prototype's communication architecture. The teacher application acts as the TCP server and receives connections from the modules used by students. For each accepted connection, an instance of the class corresponding to the assessment moment is created and executed in its own thread. In this way, students can proceed independently after connecting, while responses are forwarded to the persistence layer. The implementation uses one thread per connection, without a formal ThreadPool.

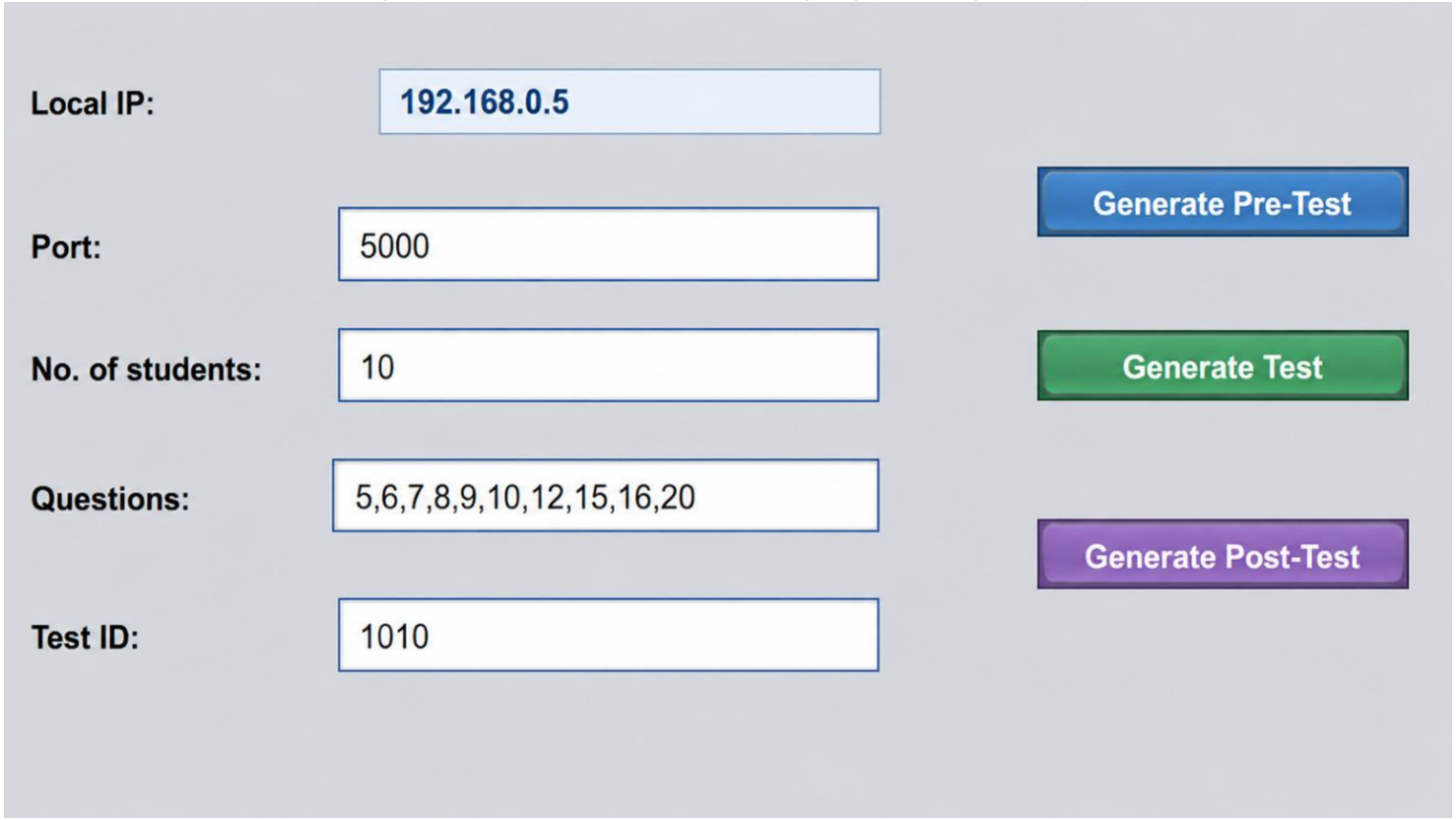
Figure B1 – TCP client-server architecture and concurrent handling by threads.



Source: Authors' own work.

Figure B2 presents the module used by the teacher to configure execution of the activity. The interface includes the server IP address and port, the expected number of students, the selected questions, and the test identifier, as well as the commands that initiate the pre-test, test, and post-test. These parameters establish the conditions required for communication with the student modules and for recording responses.

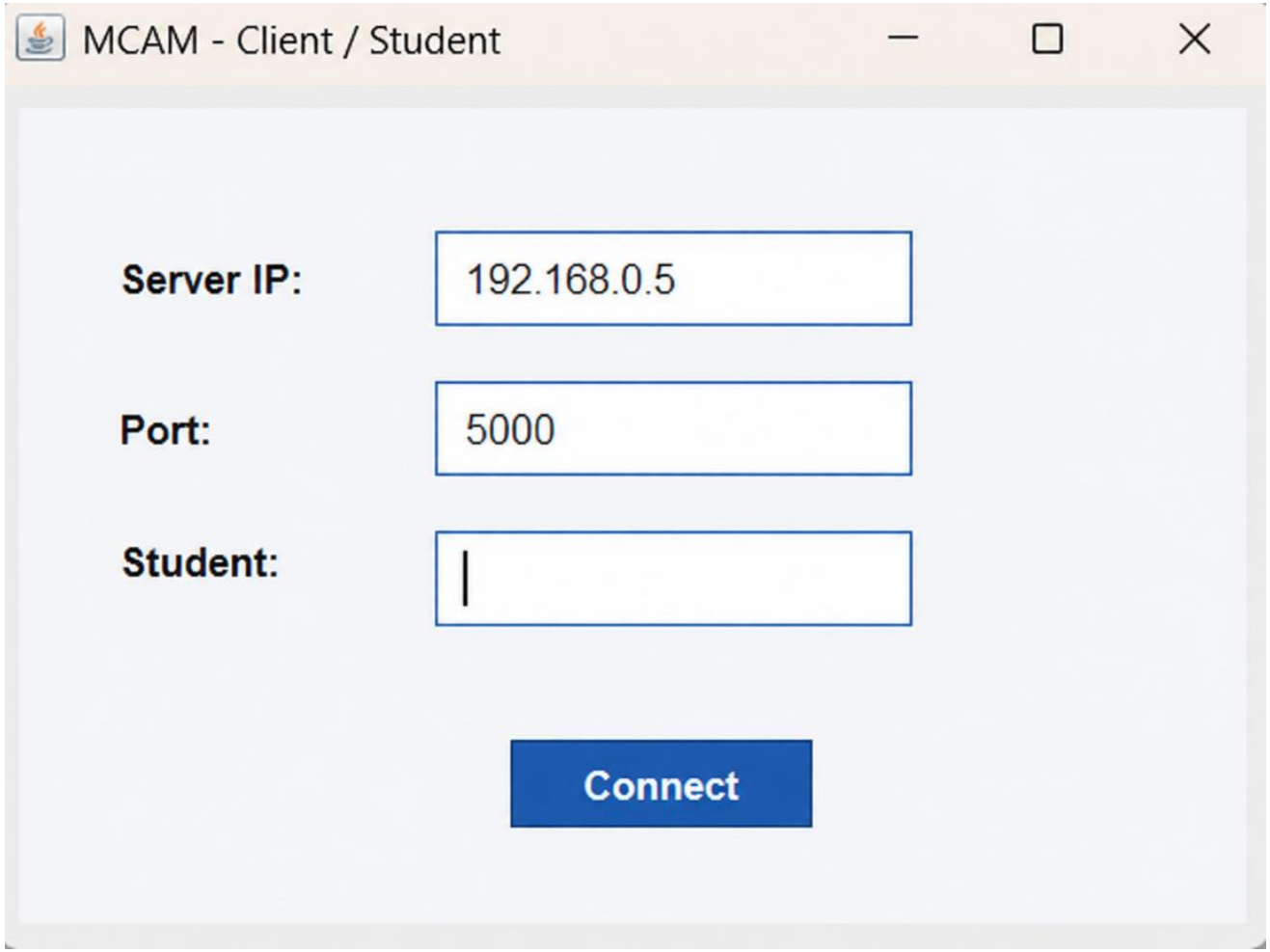
Figure B2 – Teacher module interface for configuring and running the activity



Source: Authors' own work.

Figure B3 presents the interface used by the student to establish a connection with the teacher application. The IP address and port identify the activity server, while the identification field associates the responses with the participant during execution. After connecting, the student receives the items corresponding to the assessment moment initiated by the teacher.

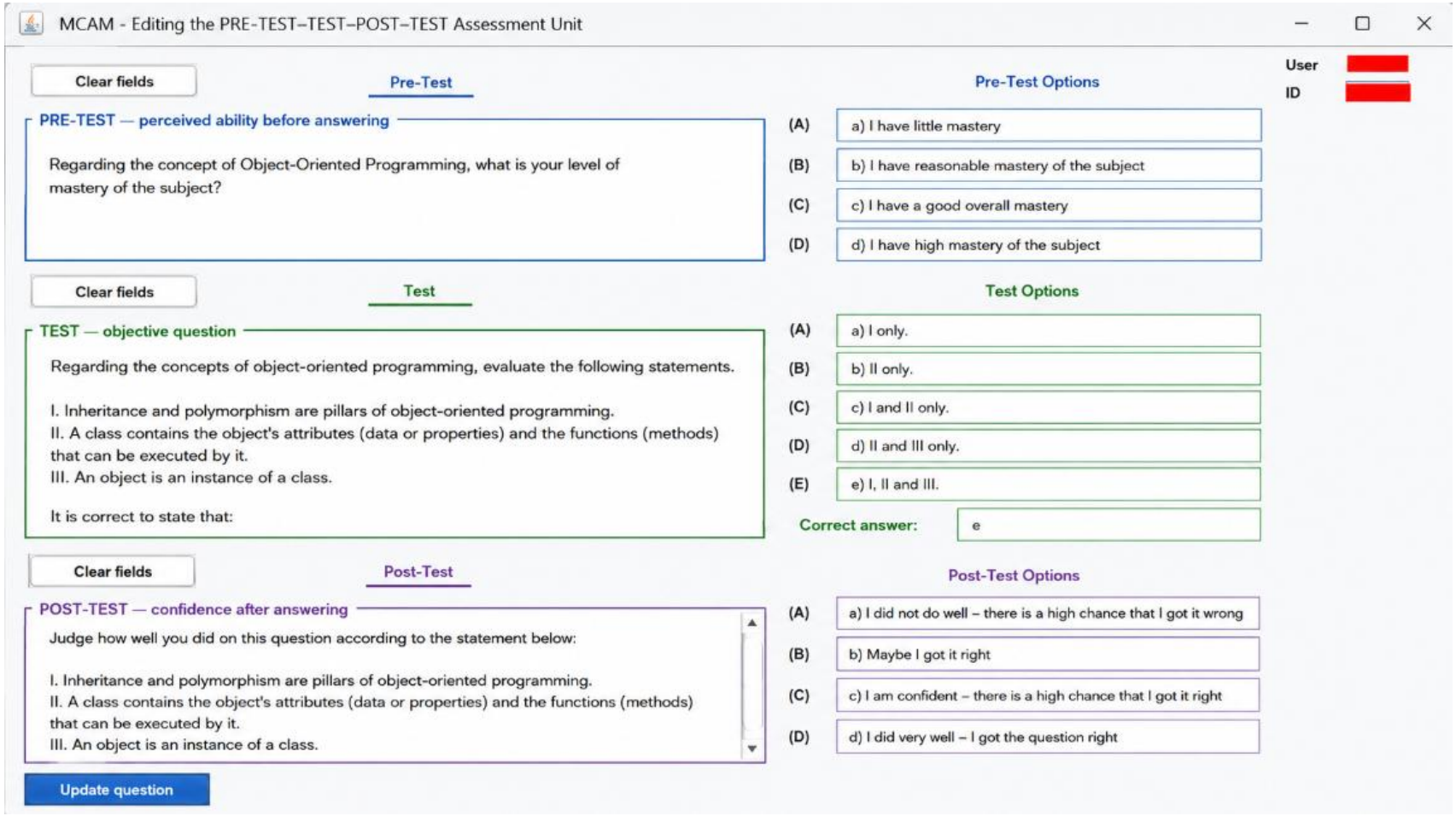
Figure B3 – Student module interface for connecting to the server



Source: Authors' own work.

Figure B4 highlights the integrated structure of the assessment unit used by the MCAM. For the same content, the teacher registers the pre-test, intended for the prospective judgment of ability, the objective test question, and the post-test, intended for the retrospective confidence judgment. Linking the three moments makes it possible to subsequently relate students' judgments to the performance demonstrated on the same set of items.

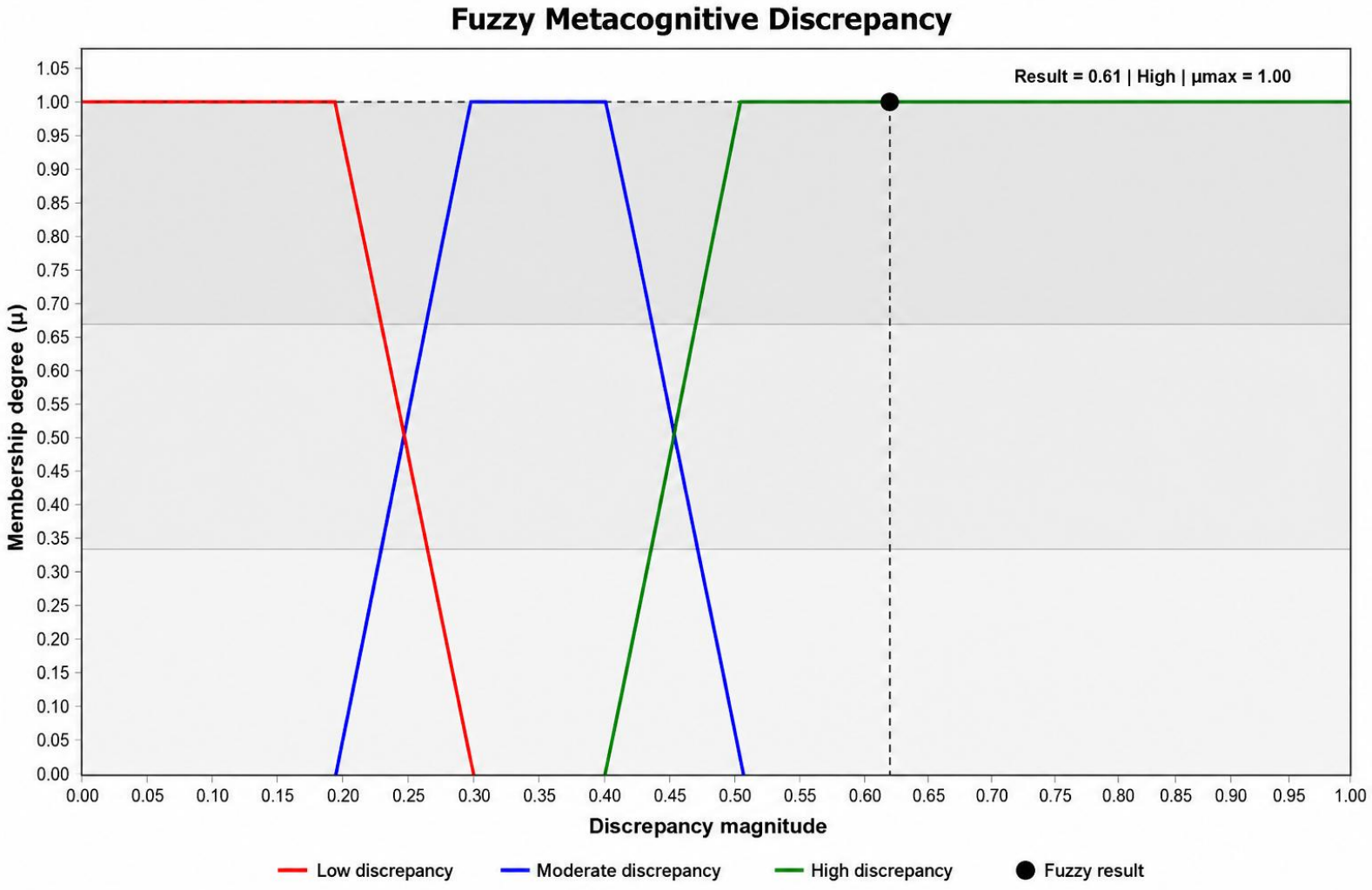
Figure B4 – Editing interface for the assessment unit composed of pre-test, test, and post-test.



Source: Authors' own work.

Figure B5 presents the output produced by the fuzzy module for one of the controlled cases in the demonstration. The curves represent membership degrees for the terms Low, Moderate, and High, and the highlighted point indicates the result obtained after defuzzification using the COG method. In the example, F ≈ 0.61 has its highest membership in the High set. This output represents the fuzzy magnitude of discrepancy; its direction remains determined by the sign of d.

Figure B5 – Fuzzy output for a controlled case of high discrepancy.

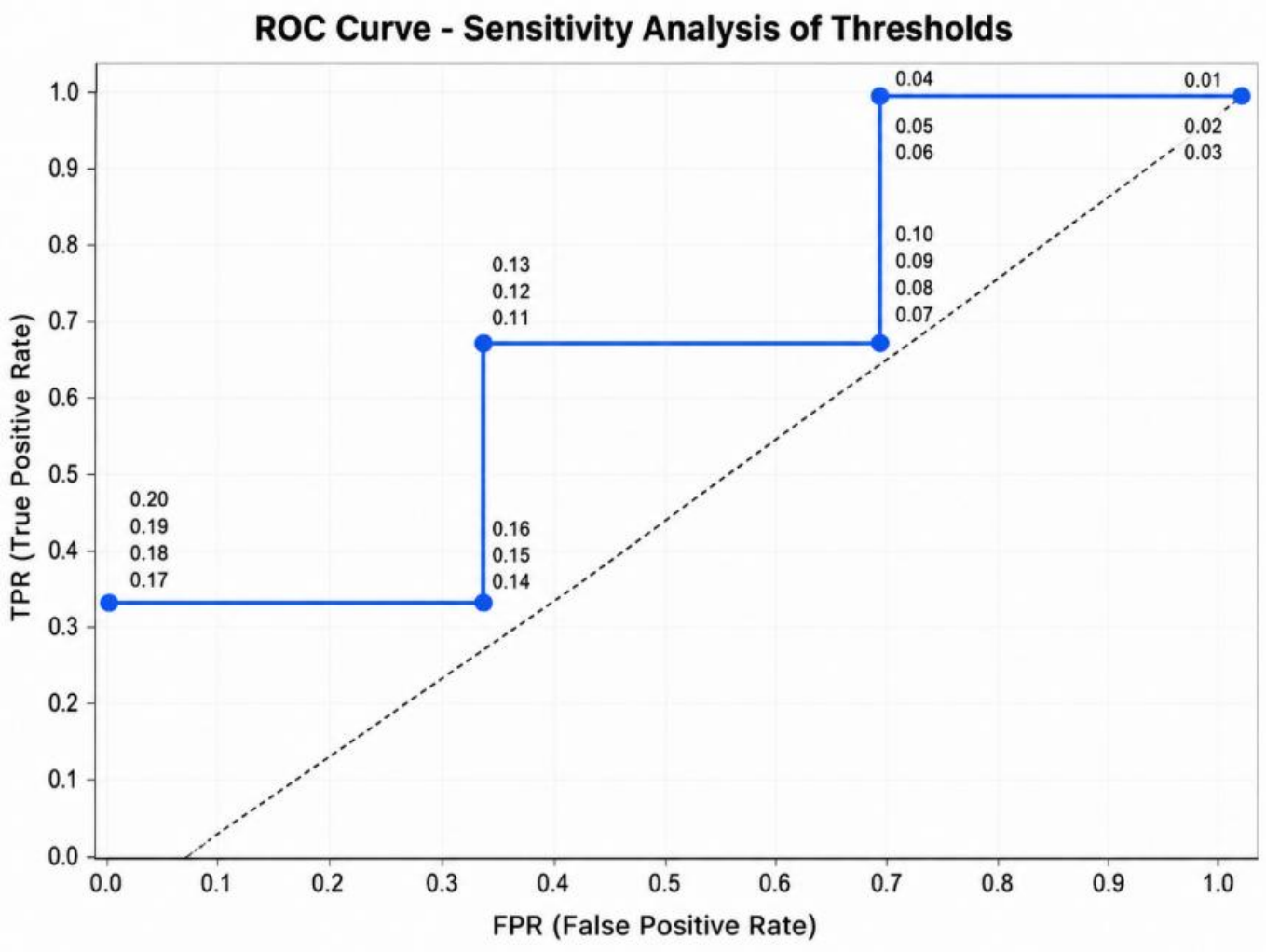


Source: Authors' own work.

Figure B6 presents the variation in TPR and FPR for the simulated cutoffs between 0.01 and 0.20. Each pair (FPR, TPR) corresponds to classification behavior at a given threshold. When different cutoffs produce the same coordinates, their labels are only displaced to preserve graph legibility. The y = x line serves

as a visual reference, allowing the sensitivity of the classification to changes in cutoff to be compared without assigning the system the task of automatically selecting an optimal threshold.

Figure B6 – ROC curve for sensitivity analysis of normalized cutoffs



Source: Authors' own work.