Giornale Italiano della Ricerca Educativa
Italian Journal of Educational Research

A structured weekly diary for metacognitive reflection in first-year engineering mathematics

Un diario settimanale strutturato come intervento metacognitivo nella matematica del primo anno di ingegneria

Maria Giulia Ballatore
Dept. of Mathematical Sciences "G.L. Lagrange", Politecnico di Torino, Torino (Italy)
Anita Tabacco
Dept. of Mathematical Sciences "G.L. Lagrange", Politecnico di Torino, Torino (Italy)

OPEN ACCESS

Double blind peer review

Citation: Ballatore, M.G., & Tabacco, A. (2026). A structured weekly diary for metacognitive reflection in first-year engineering mathematics. *Italian Journal of Educational Research*, S.I., 115-128. https://doi.org/10.7346/sird-1S2026-p115

Copyright: © 2026 Author(s). This is an open access, peer-reviewed article published by Pensa Multimedia and distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. IJEduR is the official journal of Italian Society of Educational Research (www.sird.it).

Received: June 10, 2026
Accepted: November 02, 2026
Published: November 30, 2026

Pensa MultiMedia / ISSN 2038-9744
DOI: 10.7346/sird-1S2026-p115

Abstract
First-year engineering students often struggle with university mathematics not for lack of effort, but for limited metacognitive awareness and self-regulatory skills. This single-group quasi-experimental study examines whether engagement with a structured weekly reflective diary is associated with first-session examination success in Mathematical Analysis I, and whether this association is better explained by metacognitive and affective indicators than by study time alone. The diary was implemented via Moodle across two consecutive academic years at Politecnico di Torino, in independent cohorts of 228 and 215 first-time enrolment engineering students. Diary engagement showed a replicated dose-response relationship with first-session success. Pass rates rose from 22.2% and 7.7% among non-completers to 62.6% and 78.5% among students completing 12–14 entries. Each additional entry was associated with 15% and 31% higher odds of success. The association persisted in 2025/26 after adjusting for formative test scores (OR=1.27) and was attenuated in 2024/25 (OR=1.06, p=.09). The pattern persisted without bonus points; a threshold at about eight entries emerged in 2025/26 only. Study hours did not predict grade or, after adjustment, success; negative emotional load and perceived learning were associated with success. Engagement in weeks 1–4 identified non-passers with moderate accuracy (AUC≈.67). These findings show that diary engagement is associated with academic success, alongside metacognitive and affective indicators. Causality cannot be inferred: the design cannot separate the diary from the course's bonus points and weekly communication. The diary is a low-cost, replicable intervention for first-year courses with structured weekly progressions.

Keywords: Metacognition; Reflective diaries; Mathematical Analysis; Engineering education

Riassunto
Gli studenti di ingegneria del primo anno incontrano difficoltà in matematica non per scarso impegno, ma per limitata consapevolezza metacognitiva e competenze di autoregolazione fragili. Questo studio quasi-sperimentale a gruppo singolo esamina se l'adesione a un diario riflessivo settimanale strutturato sia associata al successo nella prima sessione d'esame di Analisi Matematica I, e se tale associazione sia spiegata meglio da indicatori metacognitivi e affettivi che dal solo tempo di studio. Il diario è stato proposto su Moodle in due anni accademici consecutivi al Politecnico di Torino, a coorti indipendenti di 228 e 215 studenti alla prima iscrizione. L'adesione ha mostrato una relazione dose-risposta replicata: la percentuale di promossi è passata dal 22,2% e dal 7,7% tra chi non ha compilato alcun diario al 62,6% e al 78,5% tra chi ne ha compilati 12–14. Ogni diario in più è associato a odds di successo più alti del 15% e del 31%. L'associazione persiste nel 2025/26 dopo aver controllato per i test formativi (OR=1,27) ed è attenuata nel 2024/25 (OR=1,06, p=.09); resta anche senza punti bonus. Le ore di studio non predicono il successo; carico emotivo negativo e apprendimento percepito sì. L'adesione nelle settimane 1–4 identifica i non promossi con accuratezza moderata (AUC≈.67). Il diario è associato al successo accademico, insieme a indicatori metacognitivi e affettivi, ma non si può inferire causalità: il disegno non separa il diario dai punti bonus e dalla comunicazione settimanale. È un intervento a basso costo e replicabile nei corsi del primo anno con progressione settimanale strutturata.

Parole chiave: Metacognizione; Diari riflessivi; Analisi Matematica; Educazione ingegneristica

Credit author statement

This paper and the research behind are based on the collaborative efforts of all the authors.
MGB: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Visualization. AT: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data Curation, Writing - Review & Editing, Supervision. We have no conflict of interest to disclose. Specific funding has not been provided for this study.

1. Introduction

Teaching Mathematical Analysis I to first-year engineering students is challenging. Students moving from secondary to tertiary education often struggle with the abstract nature of concepts, the faster pace of instruction, and the shift from procedural to conceptual understanding (Ballatore & Tabacco, 2025b). These difficulties are compounded by large classes, diverse mathematical backgrounds, and varying levels of preparation (Ballatore & Tabacco, 2025a).

Beyond content mastery, learning mathematics requires metacognitive skills: the ability to plan, monitor, and evaluate one's own learning processes (Schoenfeld, 1985). Students who engage in metacognitive reflection show improved problem-solving, deeper conceptual understanding, and greater persistence (Mejares & Tan, 2023). For first-year engineering students these skills matter because university mathematics demands recognising gaps in understanding, selecting strategies, and regulating effort over time.

Reflective journaling is a promising tool for fostering metacognitive development (Borasi & Rose, 1989). In STEM and engineering education, structured reflection helps students articulate their thinking, identify knowledge gaps, and develop more effective learning strategies (Adams et al., 2003; Turns et al., 2014). Yet it remains underused in large-enrolment mathematics courses, where scalable tools for self-regulated learning are especially needed. Evidence from Italian universities shows that formative practices, such as self- and peer assessment, can foster self-regulation, but their systematic integration into teaching remains limited (Nicchia et al., 2025).

Structured diaries are particularly relevant in this context because they combine reflection with regularity, guidance, and feasibility. Unlike unstructured journals, which may lead to superficial or uneven responses, structured prompts direct students' attention to specific dimensions of learning: study practices, perceived understanding, difficulties, resource use, and emotional states. Metacognitive journals can support students in identifying strengths and weaknesses and in developing self-regulated learning strategies (O'Keeffe & Paige, 2020; Singh & Diefes-Dux, 2024; Ulson, 2024), moving them from passive exposure to active monitoring of their own learning (Borasi & Rose, 1989; Raga-Abee, 2023).

This study examines structured reflective journaling in first-year Mathematical Analysis courses at Politecnico di Torino. It builds on the DayByDay project, an active learning initiative supporting student engagement through continuous assessment and collaborative problem-solving (Ballatore & Tabacco, 2021). Despite these activities, many students continued to struggle with self-directed study: they found it difficult to identify effective strategies, recognise recurring errors, or evaluate their own preparation. This motivated the design of a structured diary as a complementary metacognitive scaffold.

The value of such an intervention may extend beyond examination outcomes. Future engineers need self-directed learning, critical reflection, and continuous improvement (Turns et al., 2014). Cultivating these habits early may support both academic success in a demanding first-year course and lifelong learning.

This study addresses three research questions: RQ1) To what extent does sustained diary engagement predict first-session examination success, and does this relationship exhibit dose-response or threshold effects? RQ2) Do metacognitive and affective indicators predict academic success independently of study time? RQ3) Can early diary engagement serve as a reliable early warning signal before failure becomes visible through summative assessment? These questions are examined across two consecutive cohorts of first-year engineering students.

The paper is organised as follows. The next section presents the theoretical framework, followed by the methodology. The results report participation profiles, the relationship between diary engagement and examination outcomes, metacognitive and affective predictors of success, resource use, and early warning indicators. The discussion and conclusion interpret these findings and outline implications and directions for future research.

2. Theoretical framework

The theoretical foundation rests on metacognition's established role in mathematical learning. Metacognition, defined as «thinking about thinking» (Flavell, 1979), encompasses both metacognitive knowledge,

what learners know about their cognitive processes, and metacognitive regulation, how they monitor and control learning activities (Schraw & Dennison, 1994). Schoenfeld's (1985) seminal work demonstrated that expert problem-solvers consistently engage in metacognitive behaviours: planning strategies, monitoring progress, and evaluating results. These self-regulatory processes distinguish successful from unsuccessful learners, particularly when confronting novel mathematical content (Garofalo & Lester, 1985).

The transition to university mathematics presents a critical juncture where metacognitive skills become essential. First-year engineering students face abstract concepts and a shift from algorithmic to conceptual understanding that requires sophisticated self-monitoring (Artigue, 2016; Gueudet, 2008). Students with strong metacognitive awareness, who accurately assess understanding, identify gaps, and adjust strategies, achieve better outcomes (Schunk & Zimmerman, 2011; Veenman et al., 2006), yet these competencies require explicit scaffolding (Kramarski & Mevarech, 2003).

Reflective writing has emerged as a powerful intervention for fostering metacognitive development (Boud et al., 2013; Moon, 1999). Articulating learning experiences externalises cognitive processes, making them available for examination (Schön, 1992). Through Vygotsky's (1978) framework of internalisation, writing transforms implicit processes into explicit, controllable knowledge—crucial in mathematics where students struggle to articulate reasoning (Mason et al., 2010).

A critical distinction concerns structure. Unstructured reflection offers flexibility but may yield superficial responses, particularly among novices lacking self-assessment frameworks (Hatton & Smith, 1995; Ward & McCotter, 2004). Structured approaches employ specific prompts directing attention toward particular learning aspects (Moon, 1999; Ryan, 2011). In mathematics education, planning prompts improve strategic thinking and monitoring questions strengthen self-assessment abilities (Kramarski & Mevarech, 2003; Mevarech & Kramarski, 1997). However, effectiveness depends on design: prompts must be specific yet flexible, targeting cognitive strategies, affective states, and self-regulatory behaviours (Tobias & Everson, 2009).

This investigation examines how engagement with a structured reflective diary relates to learning outcomes across a full semester with substantial cohorts. The weekly questionnaires prompt reflection on: cognitive (content studied), metacognitive (understanding difficulties), behavioural (study time, self-testing, resource usage), and affective (emotional states, confidence) dimensions. This multidimensional approach aligns with frameworks emphasising the interconnected nature of cognition, metacognition, and motivation (Efklides, 2011; Hannula, 2012). By linking diary data with objective performance measures, this study addresses a literature gap: empirical evidence on the relationship between structured reflective practice and actual learning outcomes in university mathematics.

3. Methodology

This study investigates the relationship between structured reflective practice and learning outcomes in Mathematical Analysis I, a mandatory first-semester course at Politecnico di Torino. The course is organised into 20 parallel sections; this research focuses on one section maintained under the same instructor across two consecutive academic years (2024/25 and 2025/26), ensuring instructional continuity.

3.1 Research design

This study adopts a single-group quasi-experimental design without a comparison condition. The structured diary was offered as an intervention to all students in the section. The two cohorts (2024/25 and 2025/26) are independent: they include different students, and the same procedure was applied in each year. Participation was voluntary, so exposure varied naturally, from no entries to all 14 weeks. The diary was part of the DayByDay package, in which formative tests and diary completion earned bonus points (see Implementation), so the design cannot separate the diary from these incentives. While precluding definitive causal claims, this design enables investigation of a critical question: under real-world conditions with student agency, what relationship exists between sustained metacognitive practice and outcomes?

The structured format addresses pedagogical and practical considerations. Predefined response options

enable completion in minutes and systematic quantitative analysis across large cohorts, while optional open text fields allow free comments. Repeated weekly measurement over 14 weeks captures the dynamic nature of learning and permits questions that cross-sectional designs cannot address: When does disengagement emerge? How do strategies evolve? What early indicators predict struggles?

3.2 Participants and context

The 2024/25 cohort comprised 228 first-time enrolled students plus 150 with examination debt (378 in total); the 2025/26 cohort included 215 new and 140 returning students (355). Overall, 733 students participated: 443 encountering the course for the first time and 290 in remedial study. Students arrived with heterogeneous preparation from diverse secondary backgrounds, primarily *liceo scientifico* and technical institutes. Analyses were restricted to first-time enrolment students for two reasons. First, 51–61% of repeating students submitted no diary entries, not as disengagement from the reflective tool but as a structural consequence of not attending the course. A distinction that cannot be operationalised independently of the predictor variable. Second, their prior exposure to course content confounds the learning and self-assessment measures at the core of the analysis. Results for the full sample including repeating students are consistent in direction but attenuated, as expected given their heterogeneity. Administrative assignment (not performance-based) suggests representativeness of the institutional population.

3.3 Implementation and data collection

The 14-week semester (late September–mid January) integrated the diary with regular instruction via Moodle. Each Sunday, the instructor communicated weekly learning tasks through Telegram and Moodle, specifying topics, problem sets, and resources. The diary questionnaire (available Sunday 12:00–Saturday 12:00) aligned directly with these communications.

Each questionnaire balanced comprehensiveness with efficiency. A context-anchored opening question echoed instructor communications. Core items followed weekly, capturing study time, self-evaluation frequency, AI tool usage, confidence ratings, and responses on difficulties and challenges. Rotating questions targeted collaborative learning, and help-seeking without excessive burden. A special winter break questionnaire (December 23–January 6) explored autonomous preparation strategies.

Concurrent formative assessments differed between cohorts: 2024/25 administered five tests (T0-prerequisite knowledge week 4; T1-T4 weeks 8, 10, 12, 14); 2025/26 used three (T1-T3 weeks 8, 12, 14). Tests were administered online without supervision, with one hour to complete them within broader windows (one week in 2024/25; 30 hours in 2025/26). Multiple-choice and numerical formats provided immediate feedback while assessing conceptual understanding and computational proficiency. Because they were unsupervised, scores may reflect help from peers or AI tools as well as individual knowledge. They are therefore treated as an indicator of engagement with the course, not as a pure measure of ability.

Summative assessment comprised the January and February examinations (first session). Later sessions are excluded because extended study would confound the comparison. Each examination had two stages: a multiple-choice test (0–15, minimum 8) giving access to a written section (0–15, minimum 8); bonus points from formative assessment, up to 3 in total, were added to the sum of the two scores if the minimum of 8 in both stages was reached. In 2024/25 the bonus followed a rubric based on five online tests; in 2025/26 the scores of three tests were summed and divided by 10. The diary added a further bonus within the same maximum of 3 points: 0 points below seven valid diaries (at least 70% completed), 0.5 points from seven, and 1 point from twelve in 2024/25 and from thirteen in 2025/26. The bonus was designed to encourage students to sit the examination and is part of the intervention package. Grades followed the Italian 0–30 scale (18 minimum passing).

3.4 Research questions

Three research questions guide the investigation.

- RQ1) To what extent does sustained diary engagement predict first-session examination success, and does this relationship exhibit dose-response or threshold effects? It is addressed by quantifying the relationship between diary engagement and first-session outcome and by testing for threshold patterns.
- RQ2) Do metacognitive and affective indicators predict academic success independently of study time? It is addressed by describing self-reported perceived learning, self-evaluation, and affective states across the semester and by testing their association with success beyond study time.
- RQ3) Can early diary engagement serve as a reliable early warning signal before failure becomes visible through summative assessment? It is addressed by estimating the discrimination of early indicators (diary engagement in weeks 1–4 and the first formative test) and its stability across cohorts.

3.5 Data analysis

Data extracted from Moodle and examination records preserved anonymised identifiers. All derived variables were computed at student level from the weekly diary responses. Diary engagement was operationalised as the total number of completed diary entries across the semester, ranging from 0 to 14. It was used both as a continuous predictor and as a five-level categorical indicator: 0, 1–3, 4–7, 8–11, and 12–14 entries. A binary indicator distinguished students who completed at least one diary from those who completed none. Early engagement, hereafter *Eng_Early*, was the number of completed entries in weeks 1–4; a dichotomous version identified students with at least three entries in this period. The maximum number of consecutive completed diaries was also used as an additional indicator of persistence.

To test whether the association depends on bonus points, we recomputed first-session success without any bonus and refitted the models on this outcome. We modelled sitting the test, and passing conditional on sitting, as separate outcomes. Candidate thresholds from 1 to 14 entries were compared with step-function logistic models using AIC. Students with no diary entries have no values for diary-derived covariates (perceived learning, study hours, affect) and were excluded from the models that include them.

Study time was the mean weekly self-reported number of hours devoted to the course across available diary entries.

Perceived learning quality was measured weekly on a 0–3 scale and averaged across completed diary entries. Perceived learning, study hours, self-evaluation, and affect were each measured with single self-report items, not with validated scales.

Self-evaluation, AI tool use, and group study were recorded weekly as binary indicators and aggregated at student level as the percentage of completed diary weeks in which the student reported self-evaluation or self-testing, using AI tools, or studying with peers.

Affective indicators were the percentage of completed diary weeks in which each emotional state was selected. Two composite indices were computed: the positive emotion index (mean of serenity, motivation, and satisfaction) and the negative emotional load index (mean of stress, anxiety, confusion, and frustration).

Learning-resource indicators were the percentage of completed diary weeks in which each resource was selected. All resources listed in the question were screened. We report those with the most consistent associations with outcomes, including YouTube, the theory textbook, and online exercises. This selection is post hoc. Academic outcomes were derived from institutional examination records. Early warning performance was evaluated as the ability to identify students who would not pass in the first session. We report the area under the ROC curve (AUC), sensitivity and specificity at the cut-off that maximises Youden's index, five-fold cross-validated AUC, and out-of-cohort validation, in which the model was fitted on one cohort and tested on the other. Test scores were expressed as a percentage of the maximum score because T1 had different scales (15 and 10 points). First-session success was coded as passing the January–February examination. Final grade was analysed only among passers, on the Italian 0–30 scale (18 minimum).

The analysis is quantitative. It employs descriptive statistics, correlation analyses, t-tests (with Welch's

correction when variances were unequal), ANOVA, chi-square tests, logistic and linear regression models, and ROC analysis, using PSPP, R, and Python. Open-text responses are not analysed in this paper. Cross-cohort validation examines whether 2024/25 patterns replicate in 2025/26 despite differences in assessment structure.

4. Results

Both cohorts showed high diary participation rates among first-time enrolment students: 84.2% in 2024/25 (N=228) and 87.9% in 2025/26 (N=215) completed at least one weekly entry.

4.1 Participation profiles

Mean completion was 8.25 entries (SD=5.29) in 2024/25 and 9.46 entries (SD=4.71) in 2025/26, out of 14. The second cohort showed both a higher participation rate and a higher mean completion, with 49.8% of students reaching the highest engagement level (12–14 entries) compared to 43.4% in the first cohort.

Completion was not uniform (Tab. 1). In both cohorts the distribution is markedly bimodal: students either completed very few diaries (15.8% and 12.1% completed none) or completed consistently (43.4% and 49.8% completed 12–14). Each intermediate level (1–3, 4–7, and 8–11 entries) accounts for less than 20% of students in both cohorts.

Level	Entries	2024/25 (N= 228)	2025/26 (N= 215)
0	0	15.8%	12.1%
1	1–3	13.6%	5.6%
2	4–7	8.3%	7.9%
3	8–11	18.9%	24.7%
4	12–14	43.4%	49.8%

Tab. 1: Diary engagement level distribution by cohort

The primary analyses focused on first-time enrolment students, since most repeating students did not attend the course and completed no diary entries. Repeating students who did engage showed outcomes comparable to first-time peers: in 2025/26 they passed the first session at 61.8%, against 56.3% for first-time students, with a similar mean number of entries (10.0 versus 9.5). This convergence is consistent with an association tied to active course attendance rather than to enrolment status.

4.2 Diary completion and examination outcome

Students who completed at least one diary entry passed the exam in the first session at substantially higher rates than non-completers: 52.6% versus 22.2% in 2024/25 ($\chi^2(1)=11.21$, $p=.001$, $\varphi=.22$), and 63.0% versus 7.7% in 2025/26 ($\chi^2(1)=28.38$, $p<.001$, $\varphi=.36$). The association was significant and consistent across both independent cohorts.

The dose-response relationship between engagement level and first-session outcome was confirmed in both cohorts (2024/25: $\chi^2(4)=24.37$, $p<.001$; 2025/26: $\chi^2(4)=58.86$, $p<.001$), as shown in Tab. 2.

		2024/25		2025/26	
Level	Entries	Pass rate	Grade	Pass rate	Grade
0	0	22.2%	24.13 (N= 8)	7.7%	19.00 (N= 2)
1	1–3	25.8%	21.63 (N= 8)	25.0%	22.33 (N= 3)
2	4–7	47.4%	22.22 (N= 9)	23.5%	20.25 (N= 4)
3	8–11	51.2%	25.05 (N= 22)	52.8%	23.29 (N= 28)
4	12–14	62.6%	26.18 (N= 62)	78.5%	23.90 (N= 84)
TOTAL		47.8%	25.14 (SD= 3.72)	56.3%	23.52 (SD= 3.51)

Tab. 2: First-session pass rates and mean final grades by diary completion level
Note. Mean final grades are computed only among students who passed in the first examination session; small cell sizes in low-engagement groups should be interpreted with caution

In 2025/26 the contrast between extremes was sharp: 7.7% of zero-entry students passed in the first session, compared to 78.5% of those at the highest level. The relationship between engagement level and final grade among passers was significant in 2024/25 (p=.001) but not in 2025/26 (p=.079), largely because almost no low-engagement students reached the first session, leaving levels 3 and 4 to account for 94% of passers.

Students who passed completed significantly more diary entries than those who did not: 10.16 versus 6.51 in 2024/25 and 11.56 versus 6.74 in 2025/26 (Welch tests, both p<.001). Among passers, diary completion correlated positively with final grade in both cohorts (2024/25: r=.373, p<.001, N=109; 2025/26: r=.202, p=.026, N=121).

Each additional completed entry was associated with 15% higher odds of first-session success in 2024/25 (OR=1.15, 95% CI [1.09, 1.22]) and 31% in 2025/26 (OR=1.31, [1.21, 1.43]). After adjusting for formative test scores, the association was attenuated in 2024/25 (OR=1.06, [0.99, 1.13], p=.09) and remained in 2025/26 (OR=1.27, [1.16, 1.39], p<.001). When adjusted for both formative tests and perceived learning, the estimates were similar (OR=1.04 and 1.29)

4.3 Sensitivity to bonus points

Without any bonus, 45.2% (2024/25) and 44.7% (2025/26) of first-time students would have passed, against the observed 47.8% and 56.3%. Six of 109 passers in 2024/25 and 25 of 121 in 2025/26 passed thanks to bonus points. The diary bonus alone changed the outcome of none and of one student, respectively. The dose-response relationship persisted on the bonus-free outcome (2024/25: $\chi^2(4)=24.58$, p<.001; 2025/26: $\chi^2(4)=48.95$, p<.001), and so did the per-entry association (Tab. 3). Engagement also predicted sitting the multiple-choice test. In 2025/26 it predicted passing among those who sat (OR=1.24), while in 2024/25 it did not (OR=1.03, p=.38). Part of the association therefore operates through reaching the examination.

Outcome	2024/25	2025/26
Pass, Unadjusted	1.15 [1.09, 1.22]	1.31 [1.21, 1.43]
Pass, Adjusted for Tests	1.06 [0.99, 1.13]	1.27 [1.16, 1.39]
Pass Without Bonus, Unadjusted	1.15 [1.09, 1.22]	1.30 [1.19, 1.42]
Pass Without Bonus, Adjusted for Tests	1.06 [0.99, 1.13]	1.26 [1.14, 1.39]
Sitting The Test, Adjusted for Tests	1.13 [1.01, 1.25]	1.32 [1.15, 1.52]
Pass Given Sitting, Adjusted for Tests	1.03 [0.96, 1.11]	1.24 [1.13, 1.36]

Tab. 3: Association between number of completed diary entries and examination outcomes, with and without bonus points
Note. Models are logistic regressions on first-time enrolled students (N=228 in 2024/25; N=215 in 2025/26). "Adjusted" models control for formative test scores. "Without bonus" means a pass defined by a multiple-choice score ≥ 8, a written score ≥ 8, and a sum ≥ 18, with no bonus points. Sitting the test and passing given sitting are modelled separately

4.4 Metacognitive indicators and study time

Mean weekly study hours, averaged across completed diary entries, were not correlated with final grade in either cohort (2024/25: r=.060, p=.519; 2025/26: r=.052, p=.573). Among first-session passers, the regression coefficient for hours, holding constant diary entries, test scores, and the other diary indicators, was negative in 2024/25 (B=−0.46, p=.026) and non-significant in 2025/26 (B=−0.09, p=.648). In 2024/25, higher reported hours were associated with lower grades among students who had already passed, consistent with the interpretation that, above a minimum threshold, additional study time signals persistent difficulty rather than productive effort.

Self-assessed learning quality showed the opposite pattern and distinguished passers from non-passers in both cohorts. Mean perceived learning across the semester was 2.07 versus 1.86 on a 0–3 scale in 2024/25 (p<.001) and 1.95 versus 1.83 in 2025/26 (p=.038). It also declined progressively in both cohorts, from 2.25 in week 1 to 1.74 in week 13 in 2024/25, and from 2.17 to 1.80 in 2025/26.

A joint model including study hours and the metacognitive and affective block clarified this contrast. Hours had no independent association with first-session success (OR per SD=1.04 and 1.05). Adding perceived learning, self-evaluation frequency, and affective indicators improved fit beyond hours (likelihood ratio test p=.004 in 2024/25 and p=.051 in 2025/26). Negative emotional load was the clearest component (OR=0.64 [0.42, 0.99] and 0.68 [0.46, 1.01]).

Among study strategies, self-evaluation distinguished passers from non-passers in 2024/25 (72.2% vs 58.2% of diary weeks, p=.009) and showed a trend in the same direction in 2025/26 (69.6% vs 60.7%, p=.085). AI tool usage showed the opposite direction in 2024/25, where non-passers reported higher AI use (5.5 weeks vs 3.6 weeks out of 14 weeks, p=.010). In 2025/26, reported AI usage was substantially higher across the entire cohort (40% in week 1 and 62% by week 7, compared to 20% and 31% in 2024/25) and the group difference was no longer significant. AI use correlated negatively with final grade among passers in both cohorts: r=−.247 (p=.013) in 2024/25 and r=−.324 (p<.001) in 2025/26, more strongly in the cohort with higher overall AI use.

Diary completion predicted final grade among first-session passers in both cohorts. Linear regression estimated the association at B=0.27 (p=.008) in 2024/25 and B=0.18 (p=.110) in 2025/26, with a cross-cohort estimate yielded B=0.20 (p=.005).

4.5 Emotional trajectory

Both cohorts began the semester with similar emotional profiles: approximately 9% of diary completers reported feeling stressed and 45% calm in week 1. By week 13, stress had risen to 49% in 2024/25 and 40% in 2025/26, and serenity had fallen to 21% and 34% respectively. Motivation declined from 76% to 56% in 2024/25 and from 70% to 54% in 2025/26. Anxiety increased from 16% to 41% and from 16% to 42%. Trajectories were parallel, but the first cohort deteriorated more, particularly in serenity. Weekly percentages refer to students who completed that week's diary. As completion declines over the semester, the composition of the group changes, so part of the trajectories may reflect selective attrition.

Negative emotional load (the composite index of stress, anxiety, confusion, and frustration) distinguished passers from non-passers (p=.002 in 2024/25 and p=.016 in 2025/26). The positive emotion index also discriminated the two groups in 2024/25 (p=.017) but not in 2025/26 (p=.130), where the gap was attenuated by the higher baseline positive affect of non-passers. Among passers, the positive emotion index correlated with final grade at nearly identical levels in both cohorts: r=.239 (p=.017) in 2024/25 and r=.240 (p=.009) in 2025/26.

4.6 Learning resources

YouTube use showed a consistent negative association with both first-session success and final grade in both cohorts. Correlations with grade among passers were r=−.245 (p<.05) in 2024/25 and r=−.253

(p<.01) in 2025/26; correlations with pass rate were r=−.282 and r=−.273 (both p<.001). No other reported resource produced a comparably consistent and replicated pattern.

Reported use of the theory textbook correlated positively with success in both cohorts (2024/25: r=.234, p=.001; 2025/26: r=.160, p=.024).

The online exercises, available through Moodle, showed a positive correlation with grade when used by 7.7% of students in 2024/25 (r=.257, p=.010), but not when institutional promotion tripled usage to 22.0% in 2025/26 (r=.063, p=.527). The predictive value of a resource thus depends on the population using it: selective voluntary adoption signals motivated engagement, whereas broad institutional access does not.

4.7 Early identification of at-risk students

Formative assessment scores decreased progressively within each cohort's schedule. In 2024/25 (maximum test score: 15), mean scores fell from 12.50 at T0 to 8.87 at T4. In 2025/26 (maximum: 10), means were 8.47, 7.15, and 7.05 for T1, T2, and T3. In both cohorts, all formative assessment scores correlated significantly with final grade and with first-session outcome.

Early diary engagement distinguished first-session passers from non-passers in both cohorts: mean entries in weeks 1–4 (Eng_Early) were 3.24 versus 2.10 in 2024/25 (p<.001) and 3.39 versus 2.18 in 2025/26 (p<.001). Dichotomising at three or more early entries produced identical discrimination: students above the threshold passed at 59.7% versus 25.3% in 2024/25 ($\chi^2(1)$=24.51, p<.001, φ=.33) and 66.0% versus 28.6% in 2025/26 ($\chi^2(1)$=23.63, p<.001, φ=.33). This is the strongest cross-cohort replication in the dataset.

Tab. 4 reports the discrimination of three early indicators for not passing in the first session. Diary engagement in weeks 1–4 (Eng_Early) discriminated moderately and consistently in both cohorts (AUC=.68, 95% CI [.61, .75] in 2024/25; AUC=.67 [.60, .74] in 2025/26). The first formative test (T1, week 8) performed well in 2024/25 (AUC=.82 [.77, .88]) and poorly in 2025/26 (AUC=.62 [.54, .70]). Adding Eng_Early improved discrimination over T1 alone in both cohorts (likelihood-ratio tests: p=.007 and p<.001), with AUC=.84 [.79, .89] and .70 [.62, .77]. In out-of-cohort validation, AUC was .68 [.60, .76] when the 2024/25 model was applied to 2025/26, and .80 [.74, .85] in the opposite direction. The combined model flagged 79% and 54% of non-passers, with specificity of 77% and 75%. The rule of fewer than three entries in weeks 1–4 flagged 35% and 26% of students (sensitivity 50% and 43%, specificity 82% and 87%). Students who skipped T1 almost never passed (100% and 91% failed).

Indicator	2024/25	2025/26
Eng_Early (weeks 1–4), AUC	.68 [.61, .75]	.67 [.60, .74]
T1 (week 8), AUC	.82 [.77, .88]	.62 [.54, .70]
Eng_Early + T1, AUC	.84 [.79, .89]	.70 [.62, .77]
Eng_Early + T1, cross-validated AUC	.83	.69
Eng_Early + T1, out-of-cohort AUC	.80 [.74, .85]	.68 [.60, .76]
Eng_Early + T1, sensitivity / specificity	.79 / .77	.54 / .75
Eng_Early + T1, PPV / NPV	.76 / .80	.60 / .70
<3 entries in weeks 1–4, sensitivity / specificity	.50 / .82	.43 / .87
<3 entries in weeks 1–4, PPV / NPV	.75 / .60	.71 / .66

Tab. 4: Discrimination of early indicators for not passing in the first session

Note. Outcome: not passing in the first session (N=209 in 2024/25; N=204 in 2025/26, students with T1). Models are logistic regressions; T1 is a percentage of the maximum score. Sensitivity and specificity refer to the Youden-optimal cut-off. Out-of-cohort AUC: each column reports a model trained on the other cohort. The rule of fewer than three entries in weeks 1–4 uses all first-time students (N=228; N=215). PPV: positive predictive value; NPV: negative predictive value

5. Discussion

The dose-response relationship between diary completion and first-session success replicated across both cohorts. In 2025/26, pass rates rose sharply at about eight completed entries, and a step model with this cutoff fitted best. In 2024/25 the increase occurred earlier, between 1–3 and 4–7 entries, and no single cutoff emerged. The threshold should therefore be read as cohort-specific and exploratory. The 2025/26 pattern is consistent with evidence on deliberate practice, where improvement is associated with a minimum regularity that allows consolidation (Ericsson et al., 1993). Sporadic reflection, like sporadic retrieval, may be insufficient to sustain the metacognitive monitoring needed across a 14-week semester.

After adjusting for formative test performance, each additional entry remained associated with higher odds of success in 2025/26 (OR=1.27), but the adjusted association was weaker and not significant in 2024/25 (OR=1.06, p=.09), although engagement still predicted sitting the test. This pattern is consistent with the diary capturing sustained attention to the weekly progression of the course. It cannot distinguish an effect of reflection from engagement with the DayByDay package as a whole.

Study hours did not predict final grade in either cohort, and among passers the coefficient for hours was negative in 2024/25 (B=–0.46, p=.026) and not significant in 2025/26 (B=–0.09, p=.648). This is consistent with the self-regulated learning framework, in which time-on-task without metacognitive guidance tends to concentrate on ineffective strategies (Zimmerman, 2002). Perceived learning quality contrasts with hours: it discriminated passers from non-passers in both cohorts. Among students who completed at least one diary, higher perceived learning was associated with higher odds of reaching the first session in 2024/25 (OR=3.29 per scale point, 95% CI [1.21, 8.91]), after controlling for test scores and diary completion. The estimate was similar in direction but not significant in 2025/26 (OR=2.32, [0.91, 5.89]). In the joint model, hours had no independent association with success once the metacognitive and affective block was included. Higher perceived learning may reflect better study allocation and the sense of efficacy needed to sustain engagement toward the exam. The diary does not measure monitoring accuracy.

Self-evaluation replicated as a behavioural discriminator of success in both cohorts. This is in line with testing effect, retrieving information from memory produces stronger retention than re-exposure to the same material (Roediger & Karpicke, 2006). AI tool use showed the opposite direction, correlating negatively with grade in both cohorts, with a stronger association in the cohort where baseline AI usage was higher (40% in 2025/26 versus 20% in 2024/25). One possible interpretation is that some students may have used AI-generated solutions as substitutes for active problem-solving, thereby reducing opportunities for the generative processes that support durable learning. The diary data cannot distinguish productive from passive AI use; studies combining diary data with AI interaction logs could clarify the mechanism.

The near-identical stress and motivation trajectories across both independent cohorts are consistent with a structural pattern linked to the course's progressive demands. Anxiety peaked during the winter break in both cohorts (50.9% and 58.5%), reflecting anticipatory distress about an imminent examination rather than current workload. Support arriving after the final lecture is therefore poorly timed. The period between week 7 and the winter break, when stress accumulates while students still attend, is the more tractable window.

Negative affect was associated with lower success in both cohorts. The direction cannot be established, because emotional load may follow difficulty as well as contribute to it. Students whose affective load increased before December were less likely to succeed in the first examination session, even when academic indicators were also considered. Positive affect correlated with final grade at nearly identical levels in both cohorts, suggesting that sustained emotional engagement is associated with learning quality among those who do reach the exam. One untested possibility is that articulating affective states in writing supports regulation, consistent with evidence that expressive writing reduces intrusive thought (Pennebaker & Smyth, 2016).

Resources that by their nature require active cognitive engagement, a structured exercise platform with automated feedback and the theory textbook, correlated positively with success when chosen deliberately by a motivated minority. When access broadened without any change in interaction quality, or when the resource was structurally passive, the correlation disappeared or reversed. These findings suggest that the relevant predictor may not be the resource itself, but the metacognitive disposition with which the student approaches it.

YouTube use illustrates the passive end of this continuum. Its negative association with both pass rate and grade was almost identical across cohorts, making it the most robustly replicated resource finding in the dataset. One possible explanation is that observing worked solutions may displace the generative retrieval processes associated with retention (Wittrock, 1992). The online exercise platform illustrates the access paradox: positively correlated with grade when few students used it in 2024/25, neutral when institutional promotion tripled adoption in 2025/26. Selective use may signal agency and initiative, whereas broad access may dilute this signal. The same logic may apply to the diary: mechanical completion and genuine reflection may lead to the same completion count while corresponding to different learning processes.

Eng_Early discriminated stably across cohorts (AUC .67–.68), whereas T1 did not (.82 versus .62). The two tests differed in length, scoring, and content, so their discrimination is not directly comparable. Fixed cut-offs did not transfer: applying the 2024/25 cut-off to 2025/26 identified only 32% of non-passers. Cut-offs therefore need local calibration, and the model is better read as a ranking tool for outreach than as a classifier. Part of the signal may reflect general disengagement, because the outcome includes students who never sat the examination.

Three findings replicated strongly across both cohorts: the negative association between YouTube use and outcomes, the early engagement discriminating power, and the progressive emotional deterioration from week 1 to the winter break. A second tier replicated in direction but not always in significance: perceived learning as a predictor, the negative hours coefficient on grade, and the diary effect on grade. The lack of significance in 2025/26 for several of these probably reflects reduced statistical power in small subgroups, particularly the near-empty low-engagement cells among passers, rather than contradictory evidence.

Several limitations should be noted. First, all diary measures are self-reported, single-item indicators rather than validated scales, and are subject to social desirability bias; students aware that diary completion carries a small grading incentive may have described their behaviour in favourable terms. The indicators describe what students reported, not their metacognitive skill or accuracy. Future work could use validated instruments to test convergent validity. Second, the study has no comparison condition, and diary completion was not randomly assigned. The association may therefore partly reflect prior differences in motivation or academic preparation that the available controls do not fully capture. The diary should be interpreted both as a pedagogical scaffold and as a behavioural trace of sustained engagement. Third, the diary was embedded in a package that also awarded bonus points for formative tests. Bonus points changed the outcome of few students (6 and 25), and the associations persisted without them. The single-group design nevertheless cannot separate the diary from the incentives and from the weekly instructor communication. Fourth, the formative tests were unsupervised, so adjusting for them controls for scores that may include external help, including from AI tools, and the adjusted estimates may therefore be incomplete. Fifth, the data come from a single institution and a single required first-year course in which mathematical problem-solving is central. Generalisation to other disciplines, other institutional settings, or voluntary diary use requires caution and empirical verification. Finally, structural differences between cohorts limit the comparability of cohort-specific effect sizes, even where direction is consistent. The early warning model was developed and validated on two cohorts of one course with cut-offs that did not transfer between cohorts.

6. Conclusions

This study asked whether engagement with a structured weekly reflective diary, implemented within a first-year Mathematical Analysis course at an Italian technical university, was associated with academic success in the first examination session, and whether this association was better explained by metacognitive and affective indicators than by study time alone. The findings support the first two research questions as associations, and the pattern held across two independent cohorts of first-time enrolment students.

Students who completed more diary entries were more likely to pass the exam in January or February, in both cohorts. The association was dose-dependent and persisted when bonus points were excluded from

the outcome. In 2025/26 it also persisted after adjusting for formative test performance. Sustained diary engagement is consistent with students remaining aligned with the course's weekly progression through the first summative assessment opportunity, whereas study hours alone did not predict success.

The pattern appears to be more closely related to self-reported metacognitive and affective indicators than to the quantity of study. Perceived learning quality, self-evaluation, and the balance between positive and negative affect were all associated with first-session success or final grade, whereas study hours were not independently associated with success. Among first-session passers, AI tool usage correlated negatively with grade in both cohorts, with a stronger association where baseline AI use was higher. What matters may be not only how much students study or which tools they access, but also how they perceive their own understanding and how actively they engage with the material.

Early diary engagement also provides a usable but modest early warning signal. Entries in weeks 1–4 identified students who did not pass with an AUC of about .67 in both cohorts, and adding the first formative test improved discrimination, with AUC between .68 and .84 across cohorts and validation sets. Performance does not replace instructor judgement. It supports flagging students for outreach before the semester's most demanding weeks.

Moreover, the outcomes of repeating students who engaged with the diary are consistent with an association tied to active course attendance rather than to enrolment status per se.

Three practical implications follow for lecturers considering similar tools. First, diary completion should be monitored as a behavioural signal in addition to formal assessment: students with few completed entries by week 9 are candidates for individual outreach well before summative results. Entries in weeks 1–4 are available by late October, and with the first formative test by mid-November, a low-cost mechanism for early intervention. Second, structured support is likely to be most useful in weeks 5–9, when the engagement gap between successful and struggling students opens and emotional load begins to accumulate; waiting until the end of the semester means intervening too late. Third, the pedagogical value of the diary depends on prompt quality. Questions asking students to identify specific misconceptions, compare self-assessment against test outcomes, or plan targeted retrieval practice are better aligned with metacognitive monitoring than generic prompts. A weekly rhythm of instructor communication specifying topics, tasks, and expected outcomes before each diary opens, appears to be a structural condition for engagement.

This study is situated in a single course at a single institution, and caution is warranted in extrapolating beyond this context. The conditions under which the diary was associated with consistent results define the scope of plausible transferability: a large mandatory first-year course with a structured weekly progression, a widely available LMS, regular instructor-student communication, and formative tests and diary completion rewarded with bonus points (3 points at most). These conditions are common in first-year STEM courses across many universities. Whether similar patterns emerge in voluntary diary contexts, upper-year courses with less structured weekly rhythms, or disciplines with different self-monitoring demands remains an open empirical question. Future research should also examine whether the metacognitive habits associated with sustained diary engagement transfer across courses over time.

Use of artificial intelligence

Artificial Intelligence tools were used only as writing and editorial assistants during the preparation of this manuscript. Specifically, ChatGPT Edu (GPT-5.5 Thinking) by OpenAI and Claude Pro (*Claude Sonnet 4.5 and Claude Sonnet 4.6*) by Anthropic were used to support language revision, title and abstract refinement, and keyword selection. AI tools were not used for data collection, data analysis, generation of results, tables, figures, or images. All AI-generated suggestions were critically reviewed, edited, and validated by the authors. The authors checked the accuracy of the content, verified consistency with the empirical analyses, and retained full responsibility for the final manuscript.

References

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design Studies*, *24*(3), 275-294. https://doi.org/10.1016/S0142-694X(02)00056-X

Artigue, M. (2016). Mathematics education research at university level: Achievements and challenges. In E. Nardi, C. Winsløw, & T. Hausberger (Eds.), *Proceedings of the First Conference of the International Network for Didactic Research in University Mathematics* (pp. 11-27). University of Montpellier and INDRUM.

Ballatore, M. G., & Tabacco, A. (2021). DayByDay: A continuous engagement for better learning in a first-year mathematical course. In H.-U. Heiß, H.-M. Järvinen, A. Mayer, & A. Schulz (Eds.), *Blended learning in engineering education: Challenging, enlightening – and lasting? Proceedings of the SEFI 49th Annual Conference* (pp. 1366-1372). SEFI. https://doi.org/10.5281/zenodo.14651470

Ballatore, M. G., & Tabacco, A. (2025a). Determinants and Academic Performance: The Role of Prerequisites in University Achievement. *Excellence and Innovation in Learning and Teaching - Open Access*, *9*(2). https://doi.org/10.3280/exioa2-2024oa19149

Ballatore, M. G., & Tabacco, A. (2025b). From deficiency to proficiency in engineering education: Gamification for sustainable learning with CIAO! course. In *Engineering Educational Games for a Sustainable Society: Play, Learn and Transform* (pp. 239-259). Cham: Springer Nature Switzerland.

Borasi, R., & Rose, B. J. (1989). Journal writing and mathematics instruction. *Educational Studies in Mathematics*, *20*(4), 347-365. https://doi.org/10.1007/BF00315606

Boud, D., Keogh, R., & Walker, D. (2013). *Reflection: Turning experience into learning*. Routledge.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6-25. https://doi.org/10.1080/00461520.2011.538645

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363-406. https://doi.org/10.1037/0033-295X.100.3.363

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16*(3), 163-176. https://doi.org/10.2307/748391

Gueudet, G. (2008). Investigating the secondary-tertiary transition. *Educational Studies in Mathematics, 67*(3), 237-254. https://doi.org/10.1007/s10649-007-9100-6

Hannula, M. S. (2012). Exploring new dimensions of mathematics-related affect: Embodied and social theories. *Research in Mathematics Education, 14*(2), 137-161. https://doi.org/10.1080/14794802.2012.694281

Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education, 11*(1), 33-49. https://doi.org/10.1016/0742-051X(94)00012-U

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281-310. https://doi.org/10.3102/00028312040001281

Mason, J., Burton, L., & Stacey, K. (2010). *Thinking mathematically* (2nd ed.). Pearson Education.

Mejares, J. A., & Tan, D. A. (2023). Journal Writing: Effects on Students' Academic Performance and Attitude towards Mathematics in Modular Distance Learning. *American Journal of Educational Research*, *11*(2), 84-90. https://doi.org/10.12691/education-11-2-7

Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2), 365-394. https://doi.org/-10.3102/00028312034002365

Moon, J.A. (1999). *Reflection in Learning and Professional Development: Theory and Practice* (1st ed.). Routledge. https://doi.org/10.4324/9780203822296

Nicchia, E., Parmigiani, D., Murgia, E., & Pario, M. (2025). Formative assessment in Higher Education: supporting student learning through selfassessment and peer assessment. *Italian Journal of Educational Research*, 86–100. https://doi.org/10.7346/sird-1S2025-p86

O'Keeffe, L., & Paige, K. (2020). Reflections on journaling: An initiative to support pre-service mathematics and science teachers. *Australian Journal of Teacher Education (Online)*, *45*(4), 76-95. https://doi.org/10.14221-/ajte.2020v45n4.6

Pennebaker, J. W., & Smyth, J. M. (2016). *Opening up by writing it down: How expressive writing improves health and eases emotional pain*. Guilford Publications.

Raga-Abee, A. (2023). Effect of Journal Writing on the Conceptual Understanding of Grade 11 Students in Mathematics. *AIDE Interdisciplinary Research Journal, 2(1)*, 327-339. https://doi.org/10.56648/aide-irj.v2i1.37

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education, 16*(1), 99-111. https://doi.org/10.1080/13562517.2010.507311

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.

Schön, D.A. (1992). *The Reflective Practitioner: How Professionals Think in Action* (1st ed.). Routledge. https://doi.org/10.4324/9781315237473

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460-475. https://doi.org/10.1006/ceps.1994.1033

Schunk, D. H., & Zimmerman, B. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. Taylor & Francis.

Singh, A., & Diefes-Dux, H. A. (2024). Student use of anchors and metacognitive strategies in reflection. In *Proceedings of the 10th Research in Engineering Education Symposium (REES 2024)*, Hubballi, India, January 4–6, 2024 (pp. 40-49). Research in Engineering Education Network. https://doi.org/10.52202/073963-0006

Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. In *Handbook of metacognition in education* (pp. 107-127). Routledge.

Turns, J. A., Sattler, B., Yasuhara, K., Borgford-Parnell, J. L., & Atman, C. J. (2014). Integrating reflection into engineering education. In *2014 ASEE Annual Conference & Exposition* (pp. 24-776). https://doi.org/10.18260/1-2—20668

Ulson, K. C. (2024). Metacognitive Journal: An Intervention Strategy in Enhancing Low-Achieving Student's Performance in Solving Functions and Perceptions Towards Mathematics. *International Journal of Humanities and Education Development (IJHED)*, *6*(3), 8-15. https://doi.org/10.22161/jhed.6.3.2

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3-14. https://doi.org/10.1007/-s11409-006-6893-0

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Ward, J. R., & McCotter, S. S. (2004). Reflection as a visible outcome for preservice teachers. *Teaching and Teacher Education, 20*(3), 243-257. https://doi.org/10.1016/j.tate.2004.02.004

Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist*, *27*(4), 531-541. https://doi.org/10.1207/s15326985ep2704_8

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2