

From Robots to Readable Learning Signals: RoboUDL for Inclusive Computational Thinking and STEM Problem-Solving in Greek Classrooms

Angelos Karabelas*

Independent Researcher, RoboVille – Robotics Center, Athens, Greece; karampelasgelos@gmail.com

*Corresponding Author: karampelasgelos@gmail.com

DOI: <https://doi.org/10.30211/JIC.202604.011>

Submitted: Jul. 01, 2026

Accepted: Sep. 09, 2026

ABSTRACT

Educational robotics can make STEM reasoning visible, but less is known about how inclusive robotics platforms can combine UDL scaffolding, learning analytics, and teacher-facing interpretation to support computational thinking and STEM problem-solving in mixed classrooms. This study designs and evaluates RoboUDL, a learning analytics-enhanced educational robotics platform for Greek classrooms. Using design science and a quasi-experimental evaluation of matched intervention and waitlist classes, the study analyzed learning outcomes and student perceptions from 288 students, 1,641 mission logs, 2,692 scaffold events, and platform evaluations from 12 dashboard-exposed educators. Findings show high implementation feasibility, reliable measurement, substantial scaffold activity, acceptable student ratings, and strong educator evaluations. STEM problem-solving showed a statistically significant comparison-adjusted gain, while computational thinking gains were positive but modest. The study contributes design knowledge for inclusive learning analytics by integrating robotics tasks, UDL-based support events, class-level aggregation, and teacher-governed dashboard signals.

Keywords: Educational robotics, Learning analytics dashboards, Universal design for learning, Computational thinking, Inclusive STEM education

1. Introduction

Educational systems are investing substantially in digital technologies, robotics, and STEM learning environments, yet the central challenge is no longer the provision of devices alone. The more consequential question is whether these technologies are embedded within accessible, pedagogically meaningful, and evidence-informed classroom systems. This issue is particularly relevant in Greece, where the 2025 Digital Decade report documents substantial investment in digital transformation alongside a severe gap in basic digital skills [1]. Greece's national roadmap includes 125 measures representing an estimated EUR 6.1 billion in public funding, demonstrating both the scale of the digital transition and the continuing need to translate investment into broadly accessible digital

capability [1]. This policy context creates a timely need for evaluated models that connect digital education resources with inclusive instructional design, classroom analytics, and measurable learning evidence.

Educational robotics provides a particularly relevant setting for addressing this challenge because robotics tasks make reasoning visible through planning, design, testing, debugging, and revision [2]. Its inclusive potential, however, cannot be assumed simply from the availability of robotics activities. The inclusive dimension of computational thinking education remains underdeveloped, with recent scoping evidence showing that relatively little is known about how programming and computational thinking can be effectively supported among children with autism spectrum disorder or attention-deficit/hyperactivity disorder [3]. This gap is consequential for mixed classrooms, where learners may vary in attention, executive function, communication preferences, sensory needs, previous experience, and confidence [3]. Inclusive educational robotics therefore requires learning environments that provide accessible entry points into tasks, visible support structures, and opportunities to participate without converting learner variability into deficit-based classifications.

Universal Design for Learning provides a design logic for addressing this requirement through learner agency and multiple means of engagement, representation, and action or expression [4]. Within educational robotics, these principles imply that learners should be able to access instructions through different representations, use support without being singled out, demonstrate understanding through different forms of response, and receive assistance during planning, sequencing, debugging, and reflection. A persistent challenge, however, concerns how these principles can be translated from broad inclusive commitments into observable and evaluable computing features. UDL is therefore most consequential for this study when it becomes embedded in the architecture of the platform through multimodal missions, optional scaffolds, alternative response pathways, and observable support interactions. This design gap motivates RQ1, which investigates how UDL principles can be operationalized within a learning analytics-enhanced educational robotics platform for inclusive STEM classrooms.

Operationalizing inclusive support also creates a second challenge concerning evidence. Teachers working in active robotics classrooms must interpret participation, incomplete tasks, repeated debugging, requests for support, and signs of difficulty while simultaneously orchestrating instruction. Learning analytics can make some of these processes more visible, but school-based dashboard research shows that analytics must address concerns involving interpretability, student privacy, teacher workload, and responsible use [5]. These concerns are especially important in inclusive classrooms because process indicators can be misinterpreted as learner classifications. A responsible dashboard should therefore organize classroom evidence for professional interpretation rather than automate judgment. The unresolved issue is how mission activity and scaffold interactions can be transformed into indicators that remain sufficiently transparent and contextualized for instructional use. Accordingly, RQ2 explores which platform-generated indicators can make participation, scaffold use, debugging, completion, and difficulty patterns visible to teachers while

preserving teacher judgment.

The learning value of such an architecture must also be established empirically. Current educational robotics evidence supports positive effects on STEM learning, but it cautions against treating STEM problem solving and computational thinking as interchangeable outcomes. Ouyang and Xu found moderate positive effects of educational robotics on STEM learning and learning attitudes while reporting that robot-assisted STEM education did not significantly improve computational thinking [2]. This distinction is important because STEM problem solving is closely connected to applied reasoning during robotics missions, whereas computational thinking represents a broader construct involving decomposition, sequencing, pattern recognition, abstraction, debugging, revision, and evaluation. The effectiveness question therefore cannot be reduced to whether students improve after participating in robotics activities. It requires determining whether an inclusive platform produces an advantage relative to an appropriate classroom comparison and whether that advantage differs across the two outcome domains. This empirical uncertainty motivates RQ3, which examines whether RoboUDL is associated with changes in computational thinking and STEM problem solving relative to a matched waitlist comparison condition.

Learning outcomes alone are also insufficient for judging an inclusive educational technology. A technically coherent platform may produce measurable process data while remaining unclear, inaccessible, difficult to use, or impractical for classroom implementation. Student experience and educator evaluation therefore represent distinct forms of evidence concerning whether the system functions as intended in authentic classroom conditions. For learners, the relevant questions concern clarity, accessibility, engagement, and confidence. For educators with direct access to the analytics environment, the relevant questions concern dashboard usefulness, inclusive feasibility, and implementation feasibility. This evaluative requirement motivates RQ4, which examines how students evaluate the robotics learning environment in terms of clarity, accessibility, engagement, and confidence, and how dashboard-exposed educators evaluate RoboUDL in terms of dashboard usefulness, inclusive feasibility, and implementation feasibility.

Consequently, these gaps reveal a broader integration problem. Prior research has examined educational robotics [2], inclusive computational thinking [3], UDL-based design [4], and learning analytics [5], but these streams are rarely integrated within a classroom-ready computing artifact that simultaneously embeds inclusive support, captures learning-process evidence, provides teacher-facing interpretation, evaluates differentiated learning outcomes, and examines user experience. RoboUDL was developed to address this integration gap as a learning analytics-enhanced educational robotics platform for inclusive computational thinking and STEM problem solving in Greek classrooms. The platform combines six robotics missions, multimodal instructions, UDL-based scaffolds, scaffold-event logging, mission process records, class- and mission-level analytics, and teacher-facing dashboard indicators. It was evaluated through a design science approach combined with a matched quasi-experimental classroom evaluation. This methodological positioning is appropriate because design science research emphasizes the development and rigorous evaluation of artifacts intended to address consequential real-world problems [6].

The contribution is threefold. At the artifact level, the study contributes a learning analytics-enhanced educational robotics platform that translates UDL into observable system features, including multimodal missions, scaffold events, process indicators, and teacher-facing interpretive rules. At the empirical level, it provides classroom evidence from Greece concerning the feasibility, accessibility, learning processes, and differentiated learning outcomes of an inclusive robotics platform in a context where digital education requires stronger classroom integration and evaluation [1]. At the design knowledge level, the study demonstrates how robotics, UDL-based scaffolding, process-event capture, class-level analytics, and teacher-governed interpretation can be integrated into a classroom-ready architecture that makes learning processes visible without converting analytics into automated learner classification.

2. Related Work and Theoretical Framing

2.1 Educational Robotics, STEM Problem-Solving, and Computational Thinking

Educational robotics research provides evidence of learning benefits, but it also reveals an important alignment problem between the learning activity, the targeted construct, and its assessment. Ouyang and Xu reported moderate positive effects on STEM learning and learning attitudes while finding no statistically significant effect on computational thinking [2]. Ching and Hsu likewise show that robotics-based computational thinking research varies in the processes emphasized and the ways those processes are assessed [7]. Consequently, these findings suggest that the educational value of robotics depends not only on exposure to robotic tasks but also on whether mission design, explicit reasoning processes, instructional support, and assessment are aligned. This distinction is central to RoboUDL because STEM problem solving is closely represented in applied mission performance, whereas computational thinking requires evidence of transferable processes such as decomposition, sequencing, pattern recognition, debugging, and evaluation. The two outcomes are therefore treated as related but analytically distinct.

2.2 Inclusive Computational Thinking and Learner Variability

Inclusive computational thinking adds a second design requirement. Oswald and colleagues show that evidence concerning programming and computational thinking for children with ASD or ADHD remains limited and methodologically heterogeneous [3]. The implication is not that inclusive robotics should be organized around diagnostic categories. Rather, platforms intended for mixed classrooms must accommodate variability without requiring students to be separated or labeled before support becomes available. This shifts the design problem from subgroup-specific accommodation toward common access, optional support, and flexible participation. It also constrains the evaluation strategy. Support categories can describe the classroom context and inform accessibility considerations, but differential effectiveness claims require larger studies specifically designed and powered for subgroup inference [3]. RoboUDL therefore treats learner variability as an architectural requirement and support-category evidence as contextual rather than diagnostic.

2.3 UDL as Platform Design Logic

UDL provides criteria for translating this architectural requirement into concrete platform functions. Recent implementation research identifies persistent challenges in maintaining explicit alignment between UDL principles and the instructional features through which those principles are enacted [8]. For an educational robotics platform, this alignment requires multiple representations of mission information, flexible ways to act and express understanding, and supports that remain available during planning, sequencing, debugging, and reflection [4], [8]. The analytical implication is equally important. Scaffold access should not be interpreted automatically as evidence of weakness. When support is intentionally available to all learners, its use can indicate help seeking, persistence, self-regulation, error detection, or teacher mediation. RoboUDL therefore treats scaffold interactions as observable process events whose meaning is interpreted alongside task completion, difficulty, debugging activity, and classroom context.

2.4 Learning Analytics Dashboards and Teacher Interpretation

Learning analytics shifts the problem from whether data can be collected to whether the resulting signals are interpretable and actionable in authentic classrooms. Possaghi and colleagues show that multimodal learning analytics can support teacher orchestration and decision making while also creating challenges involving interpretability, privacy, and workload [5]. Cabral and colleagues similarly identify continuing gaps in authentic classroom deployment and educational evaluation of learning analytics dashboards [9]. These findings imply that a useful dashboard must do more than display activity counts. Indicators need to retain instructional context, connect to transparent teacher-facing prompts, and avoid replacing professional judgment with automated classifications. This also requires distinguishing different forms of evaluation evidence. Teacher evaluations can establish perceived usefulness and implementation feasibility, whereas learning impact must be evaluated through outcome measures and process evidence. RoboUDL adopts this human-centered position by aggregating mission and scaffold information at the class and mission level before presenting it for teacher interpretation.

2.5 Design Science Integration and Research Gap

Design science provides the methodological logic for combining these requirements within one artifact and evaluating whether the resulting system addresses the identified classroom problem [6]. The synthesis produces connected design requirements. Robotics missions must elicit observable STEM and computational processes. Inclusive support must be embedded within the shared learning environment. Support and mission activity must generate interpretable process data. Those data must be transformed into transparent teacher-facing signals. Evaluation must distinguish artifact feasibility, user experience, process evidence, and learning outcomes.

RoboUDL operationalizes these requirements within a single architecture. Its distinction from conventional educational robotics platforms lies in coupling mission activity with embedded UDL support, event capture, class- and mission-level aggregation, and explicit teacher interpretation. Its distinction from standalone learning analytics dashboards lies in the fact that its indicators originate

in pedagogically structured mission and scaffold interactions rather than from a separate analytics layer added after instruction. The research gap is therefore not the absence of robotics, UDL, or dashboards individually. It is the limited integration of these elements into an evaluated computing artifact in which inclusive support and learning-process evidence are designed together. Section 3 translates this synthesis into the four-layer RoboUDL architecture.

3. RoboUDL System Design

RoboUDL was designed as a learning analytics-enhanced educational robotics platform for inclusive STEM classrooms. The system was not conceived as a repository of robotics activities. It was developed as a classroom-ready computing artifact that connects mission-based robotics tasks, UDL-based scaffolds, scaffold event logging, mission process records, dashboard indicators, and teacher-facing interpretation. This design logic reflects the need to translate inclusive principles into observable system features rather than treating accessibility as an abstract instructional value.

As shown in Figure 1, RoboUDL comprises the Student Mission Layer, Scaffold Interaction Layer, Analytics Layer, and Teacher Dashboard Layer. The four layers connect mission participation and inclusive support with event capture, class- and mission-level aggregation, and teacher interpretation. The specific functions implemented within each layer are described below.

The design of the student mission layer was guided by Universal Design for Learning. UDL emphasizes learner agency and multiple means of engagement, representation, and action or expression [4]. In RoboUDL, these principles were translated into multimodal instructions, short text, visual route maps, color-coded route segments, obstacle icons, comparison charts of expected versus actual, route alternatives, progress markers, reflection sentence starters, and oral or written response options. The design intention was to let students engage with the robotics task through multiple representational modes and demonstrate progress through multiple forms of expression.

The six-mission sequence was structured to move from basic sequencing to reflective improvement. Robot Path Basics asked students to program a route from start to finish through a clear sequence of commands. Pattern Route introduced repetition and pattern recognition through efficient movement structures. Obstacle Response required learners to plan and test a route that avoided a constraint. Debug the Route focused on error tracing by asking students to identify and correct a sequence that failed to reach the goal. Efficient Rescue Path introduced optimization by asking students to reach two targets while reducing unnecessary moves. Explain and Improve required students to revise a previous solution and justify why the improved version worked better. The sequence therefore moved from procedural command ordering toward testing, debugging, optimization, and evidence-based explanation.

The progression increased task complexity by moving from direct command ordering and pattern recognition toward constraint handling, error diagnosis, optimization, and evidence-based reflection. Students therefore moved from executing relatively bounded procedural tasks toward coordinating multiple computational and STEM reasoning processes within the same problem. This progression is consistent with educational robotics research identifying sequencing, conditional

reasoning, debugging, and algorithmic thinking as central computational thinking processes developed through robotics activities [7].

The scaffold interaction layer was designed to make support visible without treating support use as a deficit marker. Planning prompts helped students decompose the task before programming. Sequencing cues helped learners order commands and reduce working memory demands. Visual examples provided concrete representations of possible movement strategies. Debugging cues helped students compare expected and actual robot behavior. Reflection prompts supported explanation and revision. Teacher-mediated prompts captured direct teacher support when students needed clarification or strategic guidance. Scaffold use was interpreted as evidence of interaction among task demand, learner variability, and available support rather than as evidence of low ability.

The analytics layer transformed mission activity into interpretable classroom indicators. Completion rate was calculated as the number of completed attended logs divided by the total attended logs for each class and mission. Difficulty flag rate was calculated as the proportion of attended logs marked with a difficulty flag. Average scaffold access summarized the mean number of scaffold accesses across attended logs. Average debugging events summarized how often students engaged in testing and revision activity. The attendance rate summarized the exposure quality by comparing attended mission logs with assigned students. These indicators were selected because they directly relate to classroom feasibility, task accessibility, problem-solving activity, and the interpretability of learning outcome data.

The dashboard layer was designed as teacher decision support rather than automated learner classification. Completion rate, difficulty flag rate, average scaffold access, average debugging events, and attendance rate were translated into stable, monitor, or review statuses. A stable status indicated that no immediate class-level adjustment was suggested by the indicator. The monitor status indicated that the teacher should observe the pattern and consider whether clarification, pacing adjustments, or additional support were needed. A review status indicated that the teacher should examine the instructional context before interpreting the task outcome. These statuses were framed as class-level attention prompts rather than student-level diagnostic labels.

The status logic was designed to remain transparent and auditable. Stable, monitor, and review statuses were assigned at the class-mission level through predefined dashboard rules rather than through automated student profiling. Table 1 reports the indicator thresholds and status rules used to translate completion, difficulty, scaffold access, debugging, and attendance indicators into teacher-facing prompts.

Table 1. Dashboard indicator thresholds and status rules

Indicator	Threshold condition	Status	Teacher-facing prompt
Completion	At least 85 percent	Stable	Continue current instruction and consider extension activities

Completion	70 percent to below 85 percent	Monitor	Review pacing and clarify mission requirements
Completion	Below 70 percent	Review	Examine task demands, available time, and accessibility barriers before selecting additional support
Difficulty flag	Below 15 percent	Stable	Maintain current scaffolding while observing emerging patterns
Difficulty flag	15 percent to below 35 percent	Monitor	Check for recurring misconceptions and clarify instructions where needed
Difficulty flag	35 percent or higher	Review	Examine conceptual barriers, instructional clarity, and mission design
Average scaffold access	Below 1.50	Stable	Maintain scaffold availability
Average scaffold access	1.50 to 2.50	Monitor	Examine support-use patterns alongside completion and difficulty
Average scaffold access	Above 2.50	Review	Determine whether elevated support use reflects productive strategy use, unresolved difficulty, or a need for instructional clarification
Average debugging events	3.00 or fewer	Stable	Reinforce productive testing and revision
Average debugging events	Above 3.00 when the Review condition is not met	Monitor	Observe whether repeated debugging produces continued progress

Average debugging events	Above 4.00 together with a difficulty flag rate of at least 35 percent	Review	Examine blocked concepts and provide targeted debugging support
Attendance	At least 90 percent	Stable	Maintain current participation strategies
Attendance	80 percent to below 90 percent	Monitor	Examine participation patterns before interpreting mission-level indicators
Attendance	Below 80 percent	Review	Address exposure barriers before drawing class-level conclusions

Note. Thresholds are applied at the class-mission level. Statuses are teacher-facing prompts for instructional interpretation, not student-level diagnostic labels. Source: By author.

The dashboard applied five predefined class- and mission-level indicator rules. For each class mission, Review took precedence when any indicator met a Review condition. When no Review condition was present, Monitor was assigned if at least one indicator met a Monitor condition. Stable was assigned only when all indicator conditions remained within their Stable ranges. These thresholds functioned as transparent teacher attention prompts rather than validated diagnostic cutoffs.

This dashboard design reflects responsible learning analytics practice. Recent K-12 dashboard research shows that dashboards can support teacher orchestration and decision-making, but it also warns that adoption can be constrained by data interpretability, student privacy, and additional teacher workload [5]. RoboUDL responds to these concerns by keeping teacher judgment central. The dashboard does not rank students, infer disability categories, or automate instructional decisions. It organizes class-level evidence about participation, completion, difficulty, scaffold use, debugging, and attendance so that teachers can interpret classroom activity more systematically.

The platform also includes explicit interpretation rules to reduce the risk of deficit-based analytics. A low completion rate is interpreted as a prompt to examine task difficulty, time constraints, group dynamics, or accessibility barriers before focusing on individual learners. High scaffold access is interpreted as possible strategic self-regulation, productive help-seeking, or unclear task design rather than as evidence of learner weakness. High debugging events with progress are interpreted as productive computational thinking. In contrast, a combination of frequent debugging events and high difficulty flags prompts teacher review of the blocked concept or mission stage. Low attendance is treated as an exposure issue that limits the interpretability of class-level learning patterns.

As shown in Figure 2, data generation begins with student participation in a robotics mission. Mission logs record attendance, completion, time on task, difficulty, scaffold access counts, debugging events, and mission scores, while scaffold event records capture the type and timing of support use. These two record streams are aggregated at the class and mission level to produce the

indicators used for teacher interpretation. Teachers then interpret the resulting evidence through the Stable, Monitor, and Review protocol. The workflow therefore connects student activity, support interaction, analytics generation, and teacher judgment without converting process records into student-level diagnostic classifications.

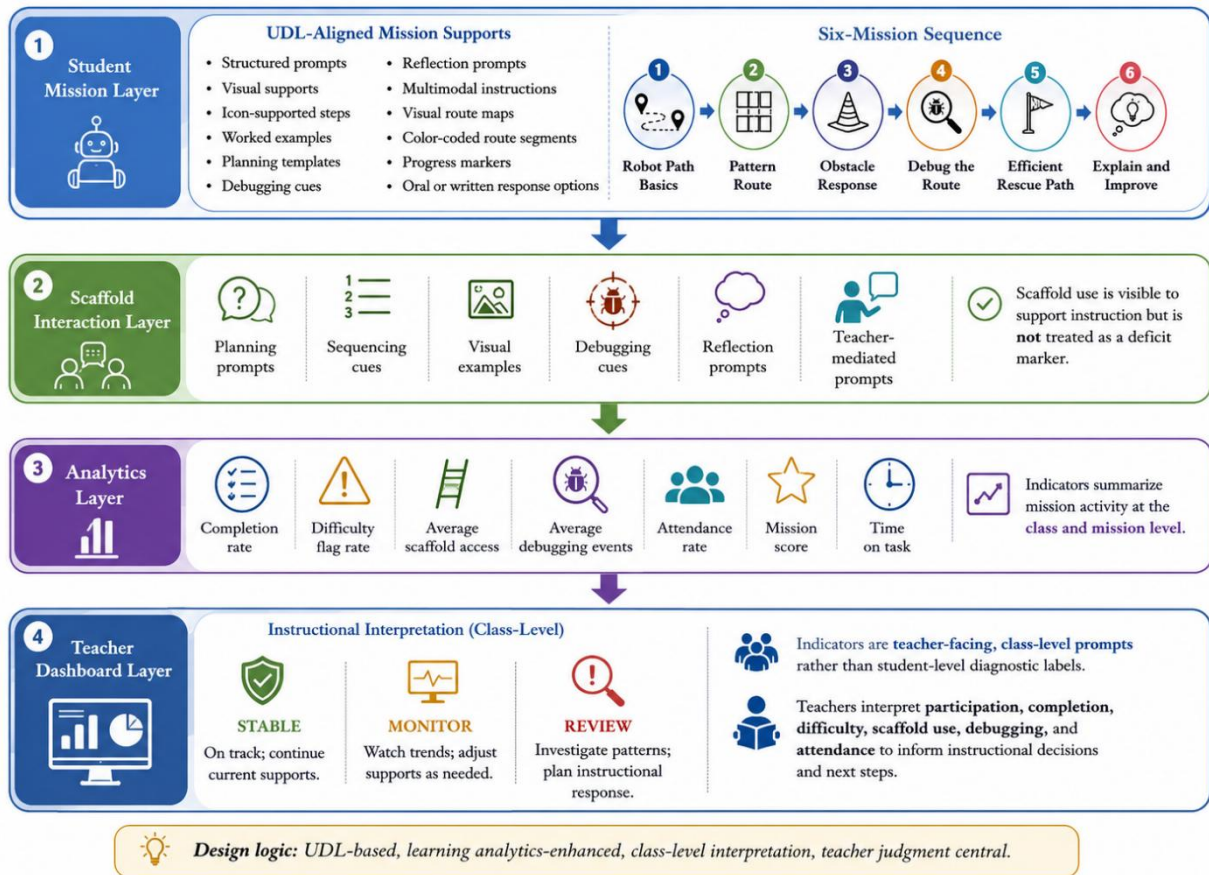


Figure 1. Four-Layer Architecture of RoboUDL

Source: By author.

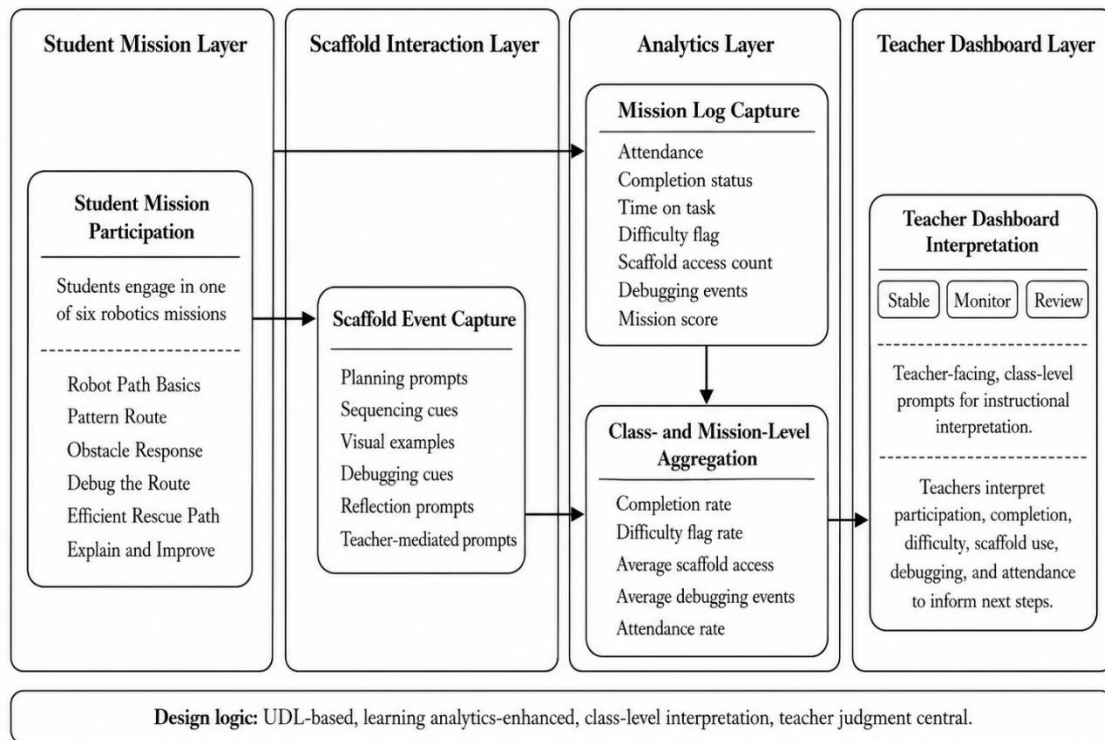


Figure 2. Mission-to-Dashboard Data Flow of RoboUDL

Source: By author.

The artifact was designed for evaluation through design science logic. Design science research emphasizes the creation and evaluation of artifacts that address relevant problems through build-and-evaluate cycles [6]. In RoboUDL, the artifact is evaluated not only on learning outcomes but also on feasibility, scaffolded activity, dashboard indicators, student perceptions, educator evaluations, and implementation fidelity. This is important because an inclusive educational technology cannot be judged solely by pre- and post-test scores. It must also be judged by whether the design is usable, interpretable, accessible, and pedagogically meaningful in classroom conditions.

4. Method

4.1 Study Design

This study used a design science research approach combined with a matched quasi-experimental classroom evaluation. Design science was appropriate because the study developed and evaluated RoboUDL as a computing artifact intended to address a practical educational problem through an implemented system, observable use, and empirical evaluation [6]. The matched classroom evaluation compared learning outcomes and student perceptions across RoboUDL intervention and matched waitlist comparison classes, while process analytics were derived from the platform records and educator platform evaluations were analyzed among the 12 educators with direct RoboUDL dashboard exposure. This design allowed the study to evaluate the platform as both a computing artifact and an instructional environment while avoiding causal claims that would require randomized assignment.

The design followed four connected phases. The first phase identified the problem of inclusive computational thinking and STEM problem-solving in Greek classrooms, where digital learning technologies require stronger evidence of meaningful classroom integration [1]. The second phase defined the platform's design objectives through educational robotics evidence, inclusive computational thinking, Universal Design for Learning, and teacher-facing learning analytics [2]-[5]. The third phase implemented RoboUDL through six robotics missions, scaffolded student supports, event logging, and dashboard indicators. The fourth phase evaluated the platform using assessment data, mission logs, scaffold events, dashboard records, student surveys, educator surveys, reliability evidence, implementation fidelity records, and documented missingness procedures.

The study was quasi-experimental because classes were not randomly assigned to conditions. Intervention and waitlist comparison classes were paired within the same school context and aligned on grade level and class size. Baseline computational thinking and STEM problem-solving scores were also examined to assess initial outcome comparability. Learner support categories were not treated as exact matching constraints and were therefore reported descriptively and used only for exploratory subgroup interpretation. This design provided a structured classroom comparison while retaining the practical conditions required for school-based implementation.

4.2 Platform Artifact

RoboUDL was the computing artifact implemented and evaluated in the intervention condition. It used the same four architecture labels shown in Figure 1. The Student Mission Layer supported the six robotics missions through multimodal task access and mission activities. The Scaffold Interaction Layer recorded optional scaffold use and teacher-mediated support events. The Analytics Layer aggregated mission logs and scaffold event records at the class and mission level. The Teacher Dashboard Layer presented the resulting indicators for instructional interpretation. Across the mission sequence, the artifact addressed sequencing, decomposition, pattern recognition, conditional reasoning, testing, debugging, optimization, evaluation, reflection, and evidence-based explanation. Learning outcomes were assessed separately through the computational thinking and STEM problem-solving measures.

4.3 Setting and Participants

The study included six anonymized schools and twelve classes in Greece. The participating classes were drawn from Grades 5 to 7, and students were 11 to 14 years old. Six classes participated in the RoboUDL intervention condition and six formed the matched waitlist comparison condition. The dataset included 288 anonymized student records (144 per condition), along with 18 educators, including classroom teachers and inclusion professionals. The implementation took place during the spring 2023 school term and involved six robotics missions delivered over six consecutive weeks. Each session lasted approximately 60 minutes.

The study was designed for mixed inclusive classrooms. Student records included broad school-confirmed support categories rather than individual diagnostic disclosures. This decision was aligned with the study's inclusive design purpose and avoided treating students with support needs as

diagnostic subgroups for strong differential effectiveness claims. Subgroup evidence was therefore treated as exploratory. The manuscript does not claim clinical or diagnostic effectiveness for ASD, ADHD, or combined support categories because several support categories were small and were not suitable for strong subgroup inference.

Baseline characteristics are summarized in Table 2. The two conditions were identical in class count, student count, grade distribution, and mean class size. Baseline computational thinking and STEM problem-solving scores differed slightly, with standardized differences of .107 and .138, respectively. Learner support categories were not used as matching constraints and are reported descriptively to make the composition of the inclusive classroom sample transparent. Comparison-adjusted gain scores were therefore used as the principal basis for interpreting learning outcomes.

Table 2a. Baseline characteristics of intervention and matched waitlist comparison conditions

Baseline characteristic	RoboUDL intervention	Matched comparison	waitlist Standardized difference	Interpretation
Number of classes	6	6	0	Balanced by design
Number of students	144	144	0	Balanced by design
Grade 5	48	48	0	Identical distribution
Grade 6	72	72	0	Identical distribution
Grade 7	24	24	0	Identical distribution
Mean class size	24	24	0	Balanced by design
Computational thinking pretest M	12.146	11.847	0.107	Small difference
STEM problem-solving pretest M	10.049	9.708	0.138	Small difference

Table 2b. Learner support composition by condition

Support category	RoboUDL intervention	Matched comparison	waitlist Interpretation
ASD support	12, 8.3%	3, 2.1%	Descriptive only
ADHD support	13, 9.0%	21, 14.6%	Descriptive

Support category	RoboUDL intervention	Matched comparison	waitlist Interpretation
			only
Combined ASD and ADHD support	11, 7.6%	3, 2.1%	Descriptive only
Other SEN support	18, 12.5%	14, 9.7%	Descriptive only
No documented support	81, 56.3%	99, 68.8%	Descriptive only
Support category not disclosed	9, 6.3%	4, 2.8%	Descriptive only

Note. Learner support distributions are reported descriptively because support category was not used as an exact matching criterion. The observed composition informs interpretation of the inclusive classroom sample but does not support condition-specific or diagnostic effectiveness claims. Source: By author.

4.4 Intervention and Comparison Conditions

The intervention classes used RoboUDL during six robotics missions. Each intervention mission included student-facing platform materials, optional scaffolds, accessibility supports, scaffold-event logging, mission-level process records, and teacher dashboard review. The intervention condition therefore combined robotics tasks with platform-supported accessibility and learning analytics.

The matched waitlist comparison classes completed standard robotics activities during the same evaluation window. These activities used comparable robotics-task time and classroom exposure but did not include the full RoboUDL platform, scaffold-event logging, or teacher-facing dashboard interpretation. Comparison classes were included to provide a matched classroom reference point for interpreting pre- and post-changes in computational thinking and STEM problem-solving. The comparison condition was not treated as a no-activity control. It represented standard robotics activity without platform-supported scaffolding, data capture, or a dashboard interpretation layer.

Both conditions completed pretest and posttest assessments using the same computational thinking and STEM problem-solving measures. The same robotics kits were used across both conditions, and class time was equivalent in duration and frequency. Different teachers delivered the intervention and comparison conditions, but all participating educators received standardized training prior to implementation. Assessment timing was identical across conditions: pretests were administered before the start of the robotics sessions, and posttests were conducted immediately after the intervention period.

4.5 Measures

A common post-implementation classroom-experience survey was administered to students in both conditions. The items assessed perceived clarity and support, accessibility, and engagement and

confidence in the robotics learning environment. Because only the intervention classes used the full RoboUDL platform, intervention-condition means are reported separately as the more direct indicator of learner experience with RoboUDL, while full-sample means provide descriptive context.

Computational thinking and STEM problem solving were assessed as related but non-equivalent outcomes because the two measures represented different forms of evidence. The computational thinking task set assessed decomposition, sequencing, pattern recognition, conditional reasoning, testing, debugging, revision, and evaluation as transferable computational processes. The STEM problem-solving rubric assessed applied performance in the robotics context through problem interpretation, STEM concept use, solution design, testing and debugging, and explanation with evidence. This operational separation prevented performance in an applied robotics problem from being treated as equivalent evidence of broader computational thinking [2], [7].

Computational thinking was measured through eight task scores, with a total score ranging from 0 to 24. The task set covered decomposition, sequencing, pattern recognition, conditions, testing, debugging, revision, and evaluation. Each task was scored from 0 to 3, and total computational thinking scores were calculated as the sum of the 8 items. Internal consistency was high for the pretest and posttest task sets, with Cronbach's alpha values of .904 and .914, respectively.

STEM problem-solving was measured using a five-dimensional rubric with a total score range of 0-20. The rubric dimensions were problem interpretation, STEM concept use, solution design, testing and debugging, and explanation with evidence. Each dimension was scored on a scale of 0 to 4. The rubric allowed students to demonstrate reasoning through plans, robot behavior, visual evidence, oral explanation, or written explanation when documented. Internal consistency was high for the pretest and posttest rubric scores, with Cronbach's alphas of .917 and .928, respectively.

Student perceptions were measured after implementation through three dimensions. Clarity and support were assessed to determine whether students could follow instructions, understand steps, use examples, and recognize when a task was complete. Accessibility was assessed by whether students could use prompts without feeling singled out, benefit from visual information, present their solution in more than one way, and experience reduced barriers. Engagement and confidence were assessed in terms of interest, retry confidence, debugging confidence, and intention to use the platform again. The reliability estimates were .875 for clarity and support, .810 for accessibility, and .826 for engagement and confidence.

Educator perceptions were measured using a 10-item post-implementation platform evaluation scale covering dashboard usefulness, inclusive feasibility, and implementation feasibility. All 18 educators participated in the study, but platform-evaluation analyses were restricted to the 12 educators with direct RoboUDL dashboard exposure, comprising six intervention teachers and six inclusion specialists who supported the intervention classes. The six comparison-only primary educators were not included in the platform-evaluation summaries because they did not receive RoboUDL dashboard exposure during the evaluation period. Internal consistency for the 10-item platform evaluation among the dashboard-exposed educators was high, with Cronbach's alpha of .892. Educator findings were interpreted descriptively as feasibility and usability evidence rather than as a

basis for inferential claims.

4.6 Assessment Scoring and Rater Agreement

Computational thinking tasks and STEM problem-solving rubrics were scored using predefined scoring guides and common scoring anchors. Two independent raters and one adjudicator completed calibration using 24 anonymized responses representing different score levels across the computational thinking task set and STEM rubric. Twenty-five percent of the STEM rubric responses were independently double-scored. Agreement for the double-scored rubric responses reached a weighted kappa of .78 and an intraclass correlation coefficient of .82. Disagreements greater than one rubric point were resolved through adjudication using the predefined scoring criteria.

4.7 Learning Analytics and Process Records

The dataset included 1,728 planned student mission sessions and 1,641 observed mission logs. Mission logs recorded attended sessions and included information on completion status, difficulty flags, scaffold access, debugging events, time on task, and mission scores. The observed mission logs were used to assess classroom feasibility, exposure, participation, completion, and mission progression.

The dataset also included 2,692 scaffold events linked to mission activity. Scaffold events included debugging cues, planning prompts, visual examples, sequencing cues, reflection prompts, and teacher-mediated prompts. These records were used to examine how the platform operationalized UDL and computational thinking support during robotics activity. Scaffold use was not interpreted as a deficit indicator. It was treated as evidence of support access, self-regulation, task difficulty, or teacher mediation, depending on the mission context.

The analytic dataset contained 72 class- and mission-level aggregate records representing the twelve classes across six missions. In the intervention condition, the 36 corresponding records were presented through the RoboUDL Teacher Dashboard Layer and used to support teacher interpretation of completion, difficulty, scaffold access, debugging, and attendance. The 36 comparison-condition aggregate records were retained for descriptive research purposes but were not presented to comparison teachers through the RoboUDL dashboard. The term dashboard-exposed record therefore refers specifically to the 36-intervention class-mission records.

4.8 Data Quality and Missingness

The dataset was checked for relational integrity, measurement quality, missingness, and reproducibility before analysis. Core joins linking students, assessments, mission records, and scaffold events contained no foreign-key violations. The imputation log documented 285 item-level filled values across the analysis files. Assessment filling affected five student records, student-survey filling affected 77 records, and educator-survey filling affected three records. Assessment values were completed through deterministic class-based and score-constrained procedures, while missing survey items were completed through item-level procedures defined within condition or educator-role groups. All filled values, affected fields, and procedures were retained in the imputation log. The main learning-outcome analysis was accompanied by a sensitivity analysis excluding records with

originally missing assessment values. This check preserved the directional outcome pattern, separating the substantive findings from dependence on the imputation procedure.

4.9 Data Analysis

The analysis followed a staged strategy. First, dataset structure, relational integrity, missingness, reliability, and implementation fidelity were examined. Second, pre- and post-changes in computational thinking and STEM problem-solving were examined for all students and by condition. Third, differences in gains were compared between the RoboUDL intervention and the matched waitlist comparison condition. Fourth, process evidence from mission logs, scaffold events, and dashboard indicators was analyzed descriptively. Fifth, student perceptions were summarized across both conditions and separately for intervention students, while educator platform evaluations were summarized for the 12 educators with direct RoboUDL dashboard exposure to assess dashboard usefulness, inclusive feasibility, and implementation feasibility.

Pre- and post-changes in computational thinking and STEM problem-solving were summarized using descriptive mean gains by condition. Between-condition gain differences were examined using independent-samples t-tests, 95% confidence intervals, and standardized mean differences using Cohen's d. Matched comparison-adjusted differences were interpreted as the main outcome evidence because they compared learning gains between the RoboUDL intervention and the matched waitlist comparison condition. Computational thinking and STEM problem-solving were interpreted separately because recent meta-analytic evidence suggests that educational robotics supports STEM learning more consistently than computational thinking development [2]. Subgroup analyses were treated as exploratory and were not used to support diagnostic effectiveness claims. Statistical analyses were conducted using SPSS version 27, and the principal descriptive, reliability, comparison-adjusted, sensitivity, and clustering calculations were independently reproduced using the analysis script accompanying the dataset. Because students were nested within classes, potential clustering was examined diagnostically using class-level intraclass correlation estimates for the gain scores. Given the twelve-class design, these estimates were used to characterize the observed magnitude of clustering rather than to support stronger causal inference.

4.10 Ethical and Data Governance Considerations

The study used anonymized classroom research data. School authorization was obtained before implementation, and participation required guardian consent, student assent, and educator consent. Student, class, school, and educator identifiers were anonymized before analysis. The dataset did not include direct identifiers, biometric data, audio, video, precise geolocation, or individual diagnostic disclosure. Support categories were broad, school-confirmed, and aggregated rather than used as individual diagnostic labels.

Data access was restricted to the research team, and the governance protocol included retention rules and transparency about dashboard interpretation. The dashboard was used only for classroom-level instructional interpretation and was not designed to profile learners, rank students, infer disability categories, or automate instructional decisions. These safeguards were important because

learning analytics in school settings requires careful attention to data interpretability, student privacy, teacher workload, and the limits of dashboard interpretation [5].

5. Results

5.1 Learning Outcomes

The computational thinking results showed positive but modest improvement. Across all students, the mean computational thinking score increased from 11.997 to 12.340 on a 0 to 24 scale, with a mean gain of 0.344. As shown in Table 3a, the RoboUDL intervention group improved from 12.146 to 12.583, while the matched waitlist comparison group improved from 11.847 to 12.097. As shown in Table 3b, the comparison-adjusted gain was 0.188, with a standardized difference of 0.218. The between-condition gain difference was not statistically significant, $t(286) = 1.846$, $p = .066$, 95 percent CI [-0.012, 0.388]. These findings indicate positive development in computational thinking, but the intervention advantage was not strong enough to support a decisive claim of effectiveness.

The STEM problem-solving results were stronger. The intervention group improved from 10.049 to 10.257 on a 0 to 20 scale, while the matched waitlist comparison group declined slightly from 9.708 to 9.660. As shown in Table 3b, the comparison-adjusted gain was 0.257, with a standardized difference of 0.324. The between-condition gain difference was statistically significant, $t(286) = 2.745$, $p = .006$, 95 percent CI [0.073, 0.441]. This finding suggests that RoboUDL was more clearly associated with applied STEM reasoning than with a separable computational thinking advantage.

Consequently, the learning outcomes show a differentiated pattern. RoboUDL produced a statistically significant comparison-adjusted advantage for STEM problem solving, whereas the computational thinking difference was positive but did not reach statistical significance. The evidence therefore supports a focused claim concerning applied STEM reasoning rather than a generalized claim of computational effectiveness. The implications of this difference in relation to prior educational robotics evidence are examined in Section 6.

Table 3a. Descriptive learning outcomes by condition

Outcome	Condition	n	Pretest M	Posttest M	Mean gain	Interpretation	
Computational thinking	RoboUDL intervention	144	12.146	12.583	0.438	Positive condition gain	within-
Computational thinking	Matched comparison	waitlist 144	11.847	12.097	0.250	Positive condition gain	within-
STEM problem-solving	RoboUDL intervention	144	10.049	10.257	0.208	Positive condition gain	within-
STEM problem-solving	Matched comparison	waitlist 144	9.708	9.660	-0.049	Slight decline	

Note. M = mean. Computational thinking scores ranged from 0 to 24. STEM problem-solving scores ranged from 0 to 20. The mean gain was calculated as the posttest mean minus the pretest mean. Source: By author.

Table 3b. Comparison-Adjusted gain results

Outcome	Comparison-adjusted gain	t(df)	p	95 percent CI	Cohen's d	Interpretation
Computational thinking	0.188	1.846(286)	.066	[-0.012, 0.388]	0.218	Positive but not statistically significant
STEM problem-solving	0.257	2.745(286)	.006	[0.073, 0.441]	0.324	Statistically significant intervention advantage

Note. CI = confidence interval. The comparison-adjusted gain is the difference between the RoboUDL intervention-group mean gain and the matched waitlist comparison-group mean gain. Cohen's d refers to the standardized between-condition difference in gain scores. Source: By author.

5.2 Platform Process Evidence

The process analytics demonstrated the platform's feasibility under classroom conditions. As shown in Table 4a, the dataset included 1,728 planned student mission sessions and 1,641 observed mission logs, indicating an overall attendance rate of 95.0 percent. Across observed mission logs, 75.3 percent of sessions were completed, 21.9 percent were partially completed, and 2.8 percent were not completed. These values indicate sustained participation across the six-mission sequence. Completion differences across conditions should be interpreted descriptively rather than as primary evidence of effectiveness because completion can be influenced by class context, task difficulty, teacher mediation, and attendance patterns.

Evidence of implementation fidelity also supported the feasibility interpretation. Across intervention classes, 36 of 36 planned RoboUDL missions were delivered as scheduled. Student-facing materials, optional scaffolds, mission logs, and dashboard records were available for each intervention mission. Teacher dashboard review was completed for all 36 intervention class mission records, providing complete teacher-facing dashboard exposure across the intervention implementation. Assessment timing was kept consistent across conditions by administering pretests before the mission sequence and posttests after completion of the implementation period. These fidelity indicators support the interpretation that the results reflect implementation of the RoboUDL platform rather than isolated or irregular robotics exposure.

The scaffold event findings connect the platform's inclusive design logic with observable learning-process data. The dataset recorded 2,692 scaffold events, including debugging cues, planning prompts, visual examples, sequencing cues, reflection prompts, and teacher-mediated prompts. The most frequent supports were debugging cues and planning prompts, suggesting that

students most often required support for error tracing, task sequencing, and strategy formation. This pattern aligns with the platform's UDL logic, which emphasizes flexible access to learning, multiple forms of representation, and multiple ways to support action and expression [4]. The distribution of scaffold events therefore provides process evidence about where support was accessed during the six-mission sequence and which forms of assistance were most frequently used during mission activity.

5.3 Dashboard Indicators and User Perceptions

The dashboard-exposed records provided empirical evidence of how the intervention classes were represented at the class-mission level. Across the 36-intervention class-mission records presented through the RoboUDL dashboard, 7 statuses were classified as stable, 14 as monitor, and 15 as review. These records represent the six intervention classes across the six-mission sequence. The wider analytic dataset contained 72 class-mission aggregate records across both conditions, but only the 36 intervention records received teacher-facing dashboard exposure. The distribution of stable, monitor, and review statuses should not be read as a simple success-or-failure profile. It shows that the dashboard generated differentiated instructional attention signals. The relatively high number of monitor and review statuses indicates that the system was sensitive to variation in completion, difficulty, scaffold use, debugging, and attendance. This supports the dashboard's role as teacher-facing learning analytics, as it helps identify where classroom interpretation or instructional adjustments may be needed.

Student perceptions provided moderate support for usability and accessibility. As shown in Table 4b, clarity and support averaged 3.445, accessibility averaged 3.518, and engagement and confidence averaged 3.553 across all students on a 1 to 5 scale. In the intervention condition, the corresponding means were slightly higher, with clarity and support at 3.483, accessibility at 3.571, and engagement and confidence at 3.595. These values should be interpreted as acceptable rather than exceptional. They indicate that students generally perceived the environment as usable, accessible, and engaging, while also suggesting room for improvement in interface clarity, scaffold personalization, and motivational design. A complete-record sensitivity checks excluding student surveys containing any filled item produced closely comparable means for clarity and support, accessibility, and engagement and confidence at 3.455, 3.521, and 3.560, respectively. The perception profile was therefore stable when analysis was restricted to fully observed survey records.

Platform evaluations among the 12 educators with direct RoboUDL dashboard exposure were stronger than the student perception ratings. Dashboard usefulness averaged 4.146, inclusive feasibility averaged 4.139, and implementation feasibility averaged 4.223 on a 1 to 5 scale. These values indicate strong perceived usefulness and feasibility across the teacher-facing dimensions of the platform. The contrast with student perceptions identifies a meaningful design priority. Students generally viewed the learning environment as usable and accessible, while the exposed educators rated the teacher-facing and implementation dimensions more strongly. Future platform refinement should therefore preserve the interpretability and practical value of the analytics architecture while improving learner-facing interface clarity, scaffold personalization, and motivational design. Educator ratings remain feasibility and usability evidence and do not independently establish learning

effectiveness [5].

Table 4a. Platform process and dashboard evidence

Evidence area	Indicator measure	or Scope denominator	or Result	Interpretation
Mission exposure	Planned student mission sessions	Full dataset	1,728	Planned exposure across the six-mission sequence
Mission exposure	Observed mission logs	Full dataset	1,641	Recorded attended mission sessions
Mission exposure	Attendance rate	1,641 of 1,728 planned sessions	95.0 percent	High implementation coverage
Mission completion	Completed sessions	Observed mission logs	75.3 percent	Most attended sessions reached completion
Mission completion	Partially completed sessions	Observed mission logs	21.9 percent	A meaningful minority required interpretation as partial completion
Mission completion	Non-completed sessions	Observed mission logs	2.8 percent	Low non-completion rate
Scaffold activity	Total scaffold events	Full dataset	2,692	Substantial use of platform-supported learning scaffolds
Scaffold activity	Most frequent scaffold categories	Scaffold event records	Debugging cues and planning prompts	Support was concentrated around error tracing, task sequencing, and strategy formation
Dashboard status	Class-mission aggregate records	12 classes × 6 missions	72	Research-level aggregate dataset across both conditions
Dashboard status	Intervention class-6 mission dashboard records	intervention classes × 6 missions	36	Dashboard-exposed instructional status records
Dashboard status	Stable status	Intervention dashboard records	7 of 36, 19.4 percent	No immediate class-level adjustment suggested
Dashboard status	Monitor status	Intervention dashboard records	14 of 36, 38.9 percent	Teacher attention to trends or support needs suggested

Evidence area	Indicator measure	or Scope denominator	or Result	Interpretation
Dashboard status	Review status	Intervention dashboard records	15 of 36, 41.7 percent	Instructional review of task context or support needs suggested

Note. The analytic dataset contained 72 class-mission aggregate records across the twelve classes and six missions. Dashboard statuses are reported for the 36 intervention records because these were the records presented to teachers through the RoboUDL dashboard. Comparison-condition aggregates were retained for descriptive research purposes and did not constitute dashboard-exposure data. Source: By author.

Table 4b. Student and educator perception evidence

Measure	Scope	Result	Interpretation
Clarity and support	All students, 1 to 5 scale	M = 3.445	Acceptable perceived clarity and support
Accessibility	All students, 1 to 5 scale	M = 3.518	Acceptable perceived accessibility
Engagement confidence	and All students, 1 to 5 scale	M = 3.553	Acceptable perceived engagement and confidence
Clarity and support	Intervention students, 1 to 5 scale	M = 3.483	Direct RoboUDL learner-experience evidence
Accessibility	Intervention students, 1 to 5 scale	M = 3.571	Direct RoboUDL learner-experience evidence
Engagement confidence	and Intervention students, 1 to 5 scale	M = 3.595	Direct RoboUDL learner-experience evidence
Educator evaluation	platform 12 exposed educators	Dashboard usefulness, M = 4.146. Inclusive feasibility, M = 4.139.	Strong perceived usefulness and feasibility

Implementation
feasibility, M = 4.223

Note. M = mean. Student and educator perception measures used a 1 to 5 scale. Full-sample student means provide descriptive context across the robotics learning environments, while intervention-student means provide the more direct evidence of learner experience with RoboUDL. Educator platform evaluations are restricted to the 12 educators with direct RoboUDL dashboard exposure. Source: By author.

Table 4c. Reliability evidence

Measure		Reliability metric	Estimate	Interpretation			
Computational thinking pretest		Cronbach's alpha	.904	Strong internal consistency			
Computational posttest	thinking	Cronbach's alpha	.914	Strong internal consistency			
STEM problem-solving pretest		Cronbach's alpha	.917	Strong internal consistency			
STEM posttest	problem-solving	Cronbach's alpha	.928	Strong internal consistency			
Clarity and support		Cronbach's alpha	.875	Strong internal consistency			
Accessibility		Cronbach's alpha	.810	Acceptable consistency	to	strong	internal
Engagement and confidence		Cronbach's alpha	.826	Acceptable consistency	to	strong	internal
Educator platform evaluation		Cronbach's alpha	.892	Strong internal consistency			

5.4 Reliability, Sensitivity, and Exploratory Subgroup Evidence

The reliability evidence supports the credibility of the measurement model. Internal consistency estimates were high for computational thinking, STEM problem-solving, student perception dimensions, and educator platform evaluation. Cronbach's alpha values ranged from .810 to .928 across the main student measures, with a value of .892 for the educator platform evaluation among the 12 dashboard-exposed educators. These values support the use of the measures for descriptive and comparison-adjusted analysis. At the same time, strong internal consistency does not remove the need for construct caution. Computational thinking remains a complex and multidimensional construct, and the modest CT results should be interpreted in light of that complexity.

Pre-post sensitivity checks examined whether the overall direction of change depended on originally missing assessment values or lower mission exposure. These checks were descriptive and did not replace the primary between-condition gain comparison reported in Table 3b. Across all analysis-ready records, mean gains were 0.344 for computational thinking and 0.080 for STEM problem solving. Excluding records with originally missing assessment values produced corresponding gains of 0.332 and 0.081. Restricting the sample to students who attended at least five missions produced gains of 0.343 and 0.071. The direction of change was therefore stable under both restrictions. Potential classroom clustering was examined separately through class-level intraclass correlation estimates, which were effectively zero for computational thinking gain and .023 for STEM problem-solving gain. Given the twelve-class design, these estimates characterize the observed magnitude of clustering rather than provide a basis for stronger causal inference.

Exploratory subgroup information was used to characterize the heterogeneity of the participating sample rather than to estimate differential effectiveness. The recorded categories comprised 15 students with ASD support, 34 with ADHD support, 14 with combined ASD and ADHD support, 32 with other SEN support, 180 with no documented support, and 13 whose support category was not disclosed. Because subgroup sizes and condition distributions were uneven, no subgroup-specific treatment effects are inferred from these categories.

Table 5. Sensitivity checks

Sensitivity sample				Outcome	n	Mean gain	Cohen's dz
All analysis-ready records				Computational thinking	288	0.344	0.397
All analysis-ready records				STEM problem solving	288	0.080	0.099
Excluding records	originally missing	assessment		Computational thinking	283	0.332	0.386

Excluding originally missing assessment records	STEM problem solving	283 0.081	0.101
At least five attended missions	Computational thinking	283 0.343	0.399
At least five attended missions	STEM problem solving	283 0.071	0.089

Source: By author.

Note. The sensitivity checks describe the stability of overall pre-post change under alternative data restrictions. Cohen's d_z denotes the paired-samples standardized mean change. The primary between-condition tests are reported separately in Table 3b.

6. Discussion

This study examined RoboUDL as a learning analytics-enhanced educational robotics platform for inclusive computational thinking and STEM problem-solving in Greek classrooms. The findings indicate that the platform's strongest contribution is not a generalized claim that robotics improves all learning outcomes. Its more precise contribution lies in designing and evaluating a computing artifact that integrates robotics missions, UDL-based scaffolding, scaffold event logging, dashboard indicators, and teacher-facing interpretation. This positioning shifts the paper from a conventional robotics intervention study to an applied computing contribution that evaluates system design, data capture, inclusive scaffolding, and teacher decision support together.

The first research question examined how UDL principles can be operationalized in a learning analytics-enhanced educational robotics platform for inclusive STEM classrooms. The findings show that UDL was operationalized through multimodal mission instructions, adjustable scaffolds, visual examples, sequencing cues, planning prompts, debugging support, reflection prompts, and alternative ways for students to demonstrate progress. This is important because UDL should not function as a rhetorical label for inclusion. It must be translated into concrete design features that reduce barriers and increase access to meaningful learning activities. CAST's UDL Guidelines emphasize learner agency and multiple means of engagement, representation, and action or expression [4]. RoboUDL extends this logic into an observable platform design by integrating scaffolding, participation, task progression, and dashboard indicators into the learning environment.

The scaffold event findings deepen the study's inclusive computing contribution by showing that scaffold use can serve as learning-process evidence. The event records indicate when support entered the mission process and which forms of assistance were used, including debugging cues, planning prompts, visual examples, sequencing cues, reflection prompts, and teacher-mediated prompts. Within a UDL-based design, a scaffold event represents an interaction among task demand, learner regulation, and available support rather than a proxy for low ability. Its interpretation therefore depends on accompanying evidence such as completion, debugging activity, difficulty, and teacher

observation. This approach extends UDL from a general accessibility framework into an observable platform architecture in which support use becomes interpretable process evidence [4], [8].

The second research question explored which platform-generated indicators capture participation, scaffold use, debugging behavior, completion, and difficulty patterns during robotics missions. The results show that the platform produced meaningful process evidence through mission logs, scaffold records, and dashboard indicators. The high attendance rate and mission-completion profile indicate that the platform was feasible in classroom conditions. The scaffold records show where students required support during the missions. The dashboard indicators translated completion, difficulty, scaffold use, debugging events, and attendance into stable, monitor, and review statuses. This finding is central to the paper's computing contribution because it shows that RoboUDL is not only a robotics activity sequence. It is a data-rich computing environment that captures learning processes and supports teacher interpretation.

The dashboard findings should be interpreted as evidence of teacher-facing learning analytics rather than as evidence of automated learner classification. This distinction matters because recent research on K-12 learning analytics dashboards shows that they can improve classroom orchestration and teacher decision-making, while also raising concerns about interpretability, student privacy, and teacher workload [5]. RoboUDL responds to this tension by using dashboard outputs as prompts for professional interpretation. The platform's value lies in organizing evidence about participation, difficulty, scaffold use, completion, debugging, and attendance so that educators can make more informed instructional judgments without ranking students, diagnosing learners, or inferring disability-specific profiles.

The third research question investigated whether RoboUDL was associated with gains in computational thinking and STEM problem-solving compared with a matched waitlist comparison condition. The findings provide stronger support for STEM problem-solving than for computational thinking. STEM problem-solving showed a statistically significant comparison-adjusted gain, while computational thinking showed a positive but modest comparison-adjusted difference that did not reach statistical significance. This contrast should be treated as an important finding rather than as an inconsistency.

This interpretation is consistent with recent evidence in educational robotics. A multilevel meta-analysis found that educational robotics produced moderate effects on STEM learning and learning attitudes, but its effect on computational thinking was minor and statistically insignificant [2]. This study follows that broader pattern while adding a platform-design explanation. RoboUDL's six-mission sequence appears to have supported applied STEM reasoning by engaging students in problem interpretation, solution planning, testing, revision, and evidence-based explanation. Computational thinking, however, is a broader and more abstract construct that may require longer exposure, more explicit instruction, repeated reflection, and more targeted assessment before stronger effects can be demonstrated.

The fourth research question examined students' evaluations of clarity, accessibility, engagement, and confidence and dashboard-exposed educators' evaluations of dashboard usefulness, inclusive

feasibility, and implementation feasibility. Student perceptions were moderate and positive, with the strongest mean observed for engagement and confidence. Among the 12 educators with direct RoboUDL dashboard exposure, dashboard usefulness averaged 4.146, inclusive feasibility averaged 4.139, and implementation feasibility averaged 4.223 on a 1 to 5 scale. The stronger educator ratings indicate that the teacher-facing analytics and implementation architecture were particularly well received. The difference between student and educator evaluations also identifies a clear development priority. Future iterations should preserve the teacher-facing value of RoboUDL while improving learner-facing interface clarity, scaffold personalization, and motivational design.

Educator findings are important, but they should not be overinterpreted. Educator perceptions support feasibility, usability, and practical relevance. They do not independently establish learning impact. This distinction is essential because acceptance data and effectiveness data answer different questions. In this study, educator ratings strengthen the argument that RoboUDL can be implemented in classrooms and can support teacher interpretation. Learning-outcome evidence must still come from the computational thinking and STEM problem-solving measures. Keeping these evidence types separate protects the manuscript from conflating perceived usefulness with measured learning development.

6.1 Contribution to Inclusive Computational Thinking and UDL-Based Design

The findings contribute to inclusive computational thinking research at the platform-design level for heterogeneous classrooms. Existing evidence concerning computational thinking and programming for children with ASD or ADHD remains limited and methodologically uneven [3]. RoboUDL addresses this problem by treating learner variability as a design requirement for the shared classroom environment rather than as the basis for disability-specific effectiveness estimates. Because several support categories were small, subgroup evidence is used to characterize participation and learning patterns rather than estimate differential treatment effects. This design orientation is relevant to classrooms in which support needs may be formal, informal, disclosed, undisclosed, temporary, or overlapping.

The support-category evidence therefore informs the inclusive design context rather than subgroup effectiveness. RoboUDL was evaluated in classrooms containing heterogeneous support profiles, but the study was not designed to estimate separate treatment effects for ASD, ADHD, combined support, or other SEN categories. The contribution lies in evaluating an inclusive platform architecture under learner variability rather than demonstrating diagnostic-group-specific effectiveness.

6.2 Implications for Digital Education Implementation

The Greek context underscores the study's practical significance. The 2025 Digital Decade report shows that Greece continues to invest substantially in digital transformation while facing a severe gap in basic digital skills [1]. Its national roadmap includes 125 measures representing an estimated EUR 6.1 billion in public funding [1]. RoboUDL addresses the corresponding classroom-level implementation challenge by demonstrating how digital education resources and robotics

activities can be integrated with inclusive design, scaffolded computational thinking tasks, dashboard-supported teacher interpretation, and classroom evaluation.

Technology provision alone does not guarantee pedagogical value. In the Greek digital education context, the broader policy challenge concerns the development of digital skills and meaningful digital transformation [1]. The classroom-level implementation challenge is whether robotics activities and digital tools are embedded in inclusive, evidence-based, teacher-usable systems. RoboUDL contributes to this transition by connecting robotics activity to task design, accessibility supports, data capture, teacher interpretation, and learning outcomes. The study is therefore relevant beyond the immediate sample because it addresses the broader challenge of moving from digital infrastructure to meaningful educational use.

6.3 Implications for Learning Analytics and Teacher Decision Support

The study also contributes to learning analytics by demonstrating a restrained, human-centered dashboard model. Many dashboard systems promise improved awareness or decision-making, but recent research warns that adoption can be limited by data interpretability, privacy concerns, and additional teacher workload [5]. RoboUDL does not treat analytics as automated decision-making. It uses analytics to organize classroom evidence and support teacher reflection. This design choice keeps human professional judgment at the center of inclusive teaching. The computing contribution lies in connecting data architecture, dashboard logic, interface design, and educational use within a teacher-facing analytics model.

The evaluation also yields design knowledge for inclusive learning analytics systems beyond the immediate RoboUDL implementation. Inclusive supports can be instrumented as process events so that support use becomes interpretable evidence of interaction with a task rather than a static accommodation or learner label. Mission and scaffold records can be aggregated at the class and mission level before translating them into teacher-facing signals, preserving instructional context while reducing pressure toward individual profiling. Dashboard statuses can remain transparent and auditable, with attention prompts under teacher judgment rather than automated classifications. These design principles connect UDL, learning analytics, and design science by showing how accessibility, data capture, and professional interpretation can be implemented within the same computing architecture [4], [5], [6], [9].

The empirical evidence defines the boundaries of this contribution. RoboUDL demonstrated high implementation coverage, meaningful scaffold activity, stable measurement, positive user evaluations, and a statistically significant comparison-adjusted advantage for STEM problem solving. Computational thinking showed a positive but smaller comparison-adjusted difference. The resulting contribution is therefore an evaluated model for connecting inclusive task design, observable learning-process evidence, and teacher-governed analytics, together with differentiated evidence about the learning outcomes that emerged during a six-mission classroom implementation.

6.4 Limitations and Future Research

The findings are bounded by several features of the study design and implementation. The design

was a matched quasi-experimental design rather than a randomized one, so causal claims must remain cautious. Different teachers delivered the intervention and comparison conditions, which means that teacher-level implementation differences cannot be fully separated from condition effects despite standardized training and matched classroom conditions. The implementation period was limited to six missions, which may have been sufficient for STEM problem-solving development but insufficient for stronger gains in computational thinking. Platform evaluation findings are descriptive because the dashboard-exposed educator sample was small. Subgroup evidence is exploratory because some support categories contained few students. Finally, the study was conducted in a specific national and classroom context, so future research should examine whether the platform produces similar implementation and learning patterns across longer deployments, different school systems, and more diverse classroom conditions.

Future research should extend RoboUDL through longer implementations, randomized or stronger quasi-experimental designs, larger inclusive samples, and more detailed analysis of scaffold trajectories. Such work should examine how specific scaffold types relate to debugging, persistence, collaboration, and transfer of computational thinking. Further studies should also investigate whether teacher dashboard use changes instructional decisions during robotics missions and whether dashboard-supported reflection improves learning over time. These directions would strengthen the evidence base while preserving the platform's central design principle. Inclusive learning analytics should support teacher judgment, learner agency, and accessible participation rather than automate classification or narrow the meaning of student progress.

7. Conclusion

This study demonstrates how inclusive educational robotics can be implemented as a teacher-governed learning analytics artifact. RoboUDL connects mission design, UDL-based supports, event capture, class- and mission-level aggregation, and dashboard interpretation so that participation, support use, debugging, difficulty, and completion become observable without being converted into learner classifications.

The classroom evaluation showed high implementation coverage, reliable measurement, meaningful scaffold activity, positive user evaluations, and a statistically significant comparison-adjusted advantage for STEM problem solving. Computational thinking development was positive but more modest. This differentiated pattern indicates that a six-mission implementation can support applied STEM reasoning while broader computational thinking development may require greater instructional duration and explicit practice.

The principal contribution is transferable design knowledge for inclusive learning analytics. Support events can function as contextual process evidence, aggregated signals can preserve instructional context, and dashboard statuses can remain transparent prompts governed by teacher judgment. Future research should test these design principles through longer deployments, larger inclusive samples, and stronger experimental or quasi-experimental designs.

Conflicts of Interest

The author declares no conflict of interest.

Funding

This research received no external funding.

References

- [1] European Commission. Greece 2025 Digital Decade Country Report. Shaping Europe's Digital Future, European Commission, 2025.
- [2] Ouyang, F. and Xu, W. The effects of educational robotics in STEM education: A multilevel meta-analysis. *International Journal of STEM Education*, 2024, 11, 7. DOI: 10.1186/s40594-024-00469-4.
- [3] Oswald, C., Paleczek, L., Maitz, K., Husny, M. and Gasteiger-Klicpera, B. Fostering computational thinking and social-emotional skills in children with ADHD and/or ASD: A scoping review. *Review Journal of Autism and Developmental Disorders*, 2024, 11, 843-862. DOI: 10.1007/s40489-023-00369-3.
- [4] CAST. CAST Universal Design for Learning Guidelines 3.0. CAST, 2024.
- [5] Possaghi, I., Vesin, B., Zhang, F., Sharma, K., Knudsen, C., Bjørkum, H. and Papavlasopoulou, S. Integrating multi-modal learning analytics dashboard in K-12 education: Insights for enhancing orchestration and teacher decision-making. *Smart Learning Environments*, 2025, 12, 53. DOI: 10.1186/s40561-025-00410-4.
- [6] Hevner, A.R., March, S.T., Park, J. and Ram, S. Design science in information systems research. *MIS Quarterly*, 2004, 28(1), 75-105. DOI: 10.2307/25148625.
- [7] Ching, Y.H. and Hsu, Y.C. Educational robotics for developing computational thinking in young learners: A systematic review. *TechTrends*, 2024, 68(3), 423-434. DOI: 10.1007/s11528-023-00841-1.
- [8] Zhang, L., Carter, R.A. Jr., Greene, J.A. and Bernacki, M.L. Unraveling challenges with the implementation of Universal Design for Learning: A systematic literature review. *Educational Psychology Review*, 2024, 36, 35. DOI: 10.1007/s10648-024-09860-7. ([Springer][1])
- [9] Cabral, L., Pinto, R. and Gonçalves, G. AI-powered learning analytics dashboards: A systematic review of applications, techniques, and research gaps. *Discover Education*, 2025, 4, 525. DOI: 10.1007/s44217-025-00964-y.

Copyright© by the authors, Licensee Intelligence Technology International Press. The article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA).