Teaching Decision Trees in Machine Learning using multiple representations

Effects on Conceptual understanding, Problem-solving performance, and Knowledge transfer ability

Oanh Tran[1]

Supervisor(s): Gosia Migut[1]**, Ilinca Rențea**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 21, 2026

Name of the student: Oanh Tran
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Ilinca Rențea, Jorge Martinez Castaneda

Abstract

Machine Learning (ML) is a rapidly growing field within Artificial Intelligence and one of the most prominent areas of technological study, which is particularly challenging for new learners as it requires a strong grasp of abstract algorithmic structures along with rigorous mathematical reasoning. Despite this, traditional instructional approaches often fail to support deep conceptual understanding, particularly for foundational models such as Decision Trees. **This study examines whether integrating multiple instructional representations (including textual explanations, visualizations, analogies, videos, and interactive simulations) enhances student learning outcomes compared to traditional text-only materials.**

To examine this lack of empirical evidence on multi-representational teaching for Decision Trees in ML education, a mixed-methods pilot experiment was employed with 10 participants was employed, comparing a multi-representation tutorial group to a text-only group. After a pre-test on mathematical and logical reasoning, participants completed a structured learning phase and a post-test measuring conceptual understanding, problem-solving, and transfer. Semi-structured interviews were also conducted to capture learner experiences.

Quantitative analysis included independent group comparisons using descriptive statistics, and inferential tests. Qualitative data were analyzed with inductive and deductive thematic analysis.

This study provides **preliminary evidence** that multi-representational instruction **may** improve problem-solving performance in Decision Tree learning. However, due to the small sample size and lack of statistical significance on two of three outcomes, these findings should be interpreted cautiously and require replication with larger samples.

The study contributes a structured evaluation framework for multi-representational ML education and provides evidence supporting the potential benefits of interactive instructional design in teaching Decision Trees.

1 Introduction

Machine Learning (ML) has become a core component of modern computer science education. As ML systems are increasingly deployed in real-world applications, there is growing demand for educational approaches that support learners in developing both theoretical understanding and practical reasoning skills.

Decision Trees are commonly used as an introductory algorithm due to their interpretability and intuitive structure [1; 2]. Despite this, students often struggle to understand key underlying concepts in Decision Trees such as purity/impurity, recursive partitioning, undefitting and overfitting [3; 4]. These difficulties arise because learners must connect mathematical representations with procedural algorithmic behavior, which is cognitively demanding.

Traditional instruction typically relies on static text and formula-based explanations. Although this method preserves the formality of mathematical definitions and calculations, it doesn't always help students truly grasp the concepts or apply what they've learned in new situations. Prior research in educational psychology suggests that multi-representational learning environments can improve comprehension by engaging multiple cognitive channels [5]. Cognitive Load Theory further suggests that poorly designed instruction may overwhelm working memory, reducing learning efficiency [6].

Despite these theoretical foundations, there is limited empirical research evaluating multi-representational instructional design specifically for Decision Trees in ML education. This represents a clear gap in the literature.

1.1 Research Question and Hypotheses

The main research question of this study is: **To what extent does multi-representational instruction improve conceptual understanding, problem-solving performance, and knowledge transfer in Decision Tree learning compared to traditional text-based materials?**

To address the research question, the following hypotheses were formulated:

- H_1: Multi-representational instruction leads to higher conceptual understanding of Decision Tree learning compared to text-based instruction amongst learners.
- H_2: Multi-representational instruction improves problem solving performance of Decision Tree learning compared to text-based instruction amongst learners.
- H_3: Multi-representational instruction results in better knowledge transfer to novel Decision Tree problems compared to text-based instruction amongst learners.

1.2 Contributions

This study presents a controlled experimental design for machine learning education research, a multi-representational Jupyter Book–based learning system, and a mixed-method evaluation framework that integrates statistical and thematic analysis. It also provides pilot empirical evidence on the effectiveness of Decision Tree pedagogy, offering initial insights into how structured, interactive instruction can support learner understanding of core machine learning concepts.

1.3 Paper structure

This paper is structured as follows: Section 2 reviews related work. Section 3 describes methodology. Section 4 presents results. Section 5 discusses findings. Section 6 addresses responsible research. Section 7 concludes the study.

2 Related Work

This section reviews theoretical foundations and empirical findings within research on multi-representational learning and machine learning (ML) education, with particular focus on Decision Tree instruction. It incorporates findings from cognitive science, mathematics education, computer science

(CS) education, and emerging ML pedagogy to identify unresolved questions regarding how instructional representations influence conceptual learning.

2.1 Foundations of Multi-representational Learning

The use of multiple external representations (MERs) to support learning is grounded in several established cognitive theories. Mayer's Cognitive Theory of Multimedia Learning deems that learners benefit when information is presented through both verbal and visual channels, as these are processed in separate cognitive subsystems, and therefore expanding effective working memory capacity [5]. This theory focuses on ideas like spatial and temporal connection, suggesting related information should appear at the same time and close together so that unnecessary mental effort is reduced. Similarly, Dual Coding Theory argues that combining visual and verbal representations creates two distinct memory traces, improving retention and understanding [7].

However, integrating MERs are not always affective. Cognitive Load Theory points out that working memory has limited capacity, so when learners must combine information from multiple sources on their own, their attention can become divided and this increases unnecessary mental effort and can make learning more difficult [6]. For example, when a diagram and its text are shown separately, learners must use mental effort to match the correct parts. This reduces the resources available for building understanding. In machine learning education, this suggests that adding more formats like visualizations, analogies, or interactive simulations can be harmful if they are not well connected.

2.2 Multiple representations in Mathematics and CS Education

Prior research in Mathematics and CS education has demonstrated that MERs can enhance learning, but the evidence is nuanced. In mathematics, research shows that combining diagrams directly with solution steps, instead of showing them separately, significantly improves students' performance in geometry problem solving [8]. Similarly, in CS education, visualizations of algorithms have been shown to improve conceptual understanding, particularly for novice learners [7]. These findings support the use of integrated, well-sequenced MERs.

Research on visual and interactive tools for teaching ML suggests mixed outcomes. Visual programming environments (e.g., Orange, LearningML) can increase engagement and make ML concepts more accessible [9; 10]. However, not all research finds positive results. Some studies show that too many representations at once can overload learners' cognitive capacity, this can also lead them to focus on surface details instead of the deeper structure of the material [11]. The order in which representations are introduced also matters. The concreteness fading approach starts with concrete, hands-on examples and gradually moves toward more abstract and symbolic forms. This has been effective in mathematics, but its benefits in CS education are ambiguous [11; 12].

2.3 Multiple representations in Decision Trees

Decision Trees are widely used as introductory ML models because of their relative interpretability and structural similarity to human decision-making processes [13]. Their branching structure provides a natural bridge between intuitive reasoning and formal classification, making them common in K–12 and introductory ML curricula [14; 9]. Despite this, existing research has predominantly focused on implementing Decision Tree algorithms in code or using simplified tools (e.g., spreadsheets, unplugged activities), rather than on systematically investigating how different instructional representations impact learning outcomes [15; 16]. These approaches generally assess whether students can build or execute Decision Trees but rarely investigate how students conceptually interpret the underlying representational structure.

Recent work by Fleischer and Biehler [3] addresses this gap by examining how students construct data-based Decision Trees after an introductory ML teaching unit. Their study moves beyond performance-based evaluation to analyze students' conceptual reasoning during tree construction. They found that students frequently struggled with selecting meaningful splitting attributes, interpreting branch semantics, and connecting tabular data to hierarchical tree structures. Importantly, students often treated decision trees as procedural artifacts rather than conceptual models of data partitioning. This finding is highly relevant to the present study because it suggests that learning difficulties in Decision Trees are fundamentally representational. Students must coordinate multiple abstractions simultaneously: raw tabular data, feature comparisons, branching logic, and classification outcomes. Difficulty in translating between these representations may partially explain why conceptual misunderstandings persist even when students can mechanically construct trees. Supporting this interpretation, Podworny et al. [4] observed that young learners could engage productively with data-driven tree construction but required substantial scaffolding to connect data patterns with branching decisions. Their work suggests that interpretability alone does not guarantee conceptual accessibility; students still require explicit representational support.

However, the literature also presents conflicting evidence regarding instructional interventions. Visual and interactive ML tools are often assumed to improve learning due to increased transparency, yet some studies report that richer interfaces can inadvertently distract learners or encourage shallow pattern matching [11]. In such cases, added representations may amplify rather than reduce cognitive complexity. Thus, while Decision Trees appear well suited to multi-representational instruction, empirical evidence remains insufficient to determine whether combining text, visuals, analogies, simulations, and videos improves conceptual learning or increases overload.

2.4 Research Gap and Contribution of this study

In summary, while the theoretical and empirical literature on multi-representational learning is rich, its application to ML education, and specifically to teaching Decision Trees remains underexplored. Prior studies have either focused on other domains (mathematics, physics, CS algorithms), used different ML topics (neural networks, classification basics), or lacked controlled experimental designs. Moreover, no prior work has systematically compared an integrated, multi-representational approach (combining text, visualizations, analogies, videos, and interactive simulations) against a traditional text-only condition for teaching Decision Trees, while also collecting qualitative data on learner experiences.

This study addresses this gap by combining a controlled experiment with qualitative analysis to evaluate whether multi-representational instruction enhances learning outcomes and to explore potential drawbacks, such as cognitive overload or distraction. By doing so, it is aimed to provide evidence-based guidance for educators and curriculum designers in ML education. However, given the relatively small sample size, the findings should be interpreted with caution and may not be sufficient to draw broadly generalizable conclusions.

3 Methodology

This study employed a mixed-methods experimental design to investigate the effect of multi-representational instructional materials on student understanding of Decision Trees in ML education. The design was chosen to enable both quantitative measurement of learning outcomes and qualitative exploration of learners' experiences. Participants were divided into two groups, receiving either a multi-representational instructional Decision Tree tutorial or a text-based learning tutorial. Learning gains were assessed using a pre-test and post-test framework, while additional qualitative data were collected through interviews. Together, these methods allowed for a comprehensive evaluation of both performance differences and subjective learning experiences, forming the basis for subsequent statistical and thematic analysis.

3.1 Conceptual background of Decision Trees

A Decision Tree is a Supervised ML algorithm used for both classification and regression tasks. It models decision-making as a hierarchical tree structure consisting of a root node, internal decision nodes, branches, and leaf nodes. Each internal node represents a question on a feature, each branch corresponds to the outcome of that question, and each leaf node represents a final prediction or class label. Decision Trees recursively split data based on feature values to minimize impurity.

To effectively learn and apply Decision Trees, students should develop an understanding of several fundamental concepts:

1. **Tree Structure and Recursive Splitting**
   - Data are repeatedly partitioned into smaller subsets based on feature values.
   - The roles of root nodes, internal nodes, branches, and leaf nodes.
2. **Entropy and Information Gain**
   - Entropy measures the uncertainty or disorder within a dataset.
   - Information Gain quantifies how much uncertainty is reduced after a split.
3. **Gini Impurity**
   - An alternative measure of node impurity commonly used in classification trees.
   - In Decision Trees, splits are chosen to minimize the weighted Gini Impurity of child nodes.
4. **Overfitting and Underfitting**
   - Deep trees can memorize training data, leading to poor performance on unseen data.
   - Decision tree models can be too simple to capture the underlying patterns in the dataset

These concepts could be challenging for new learners because they require mathematical reasoning, understanding probability, step-by-step computational logic, and the ability to visualize hierarchical structures.

3.2 Learning objectives

The learning objectives in this tutorial are based on Bloom's Revised Taxonomy and cover the first four cognitive levels: **Remember, Understand, Apply, and Analyze** [17]. This ensures a clear progression from basic knowledge of Decision Trees to applying and analyzing their behavior in simple cases. The higher levels, Evaluate and Create, are not included due to the limited scope and time constraints of the tutorial and this research.

Bloom Level	Learning Objectives
Remember	Identify the key components of a Decision Tree classifier, including root nodes, internal nodes, branches, and leaf nodes.
Understand	Explain the purpose of Decision Tree classifiers in supervised machine learning and describe how they partition data into decision regions.
Understand	Explain the recursive process of Decision Tree construction and how datasets are split into progressively smaller subsets.
Understand	Explain the concept of Gini impurity as a splitting criterion in Decision Trees.
Understand	Explain Entropy and Information gain aas a splitting criterion in Decision Trees.
Apply	Calculate Gini impurity for a given dataset and interpret what the value indicates about class purity.
Apply	Calculate Entropy and Information gain for a dataset and use the results to compare potential splits.
Analyze	Compare the differences between the 2 splitting criteria (Gini impurity vs. Entropy and Information gain)

Table 1: Learning objectives structured according to Bloom's Revised Taxonomy, used for designing the instructional materials and post-test questions

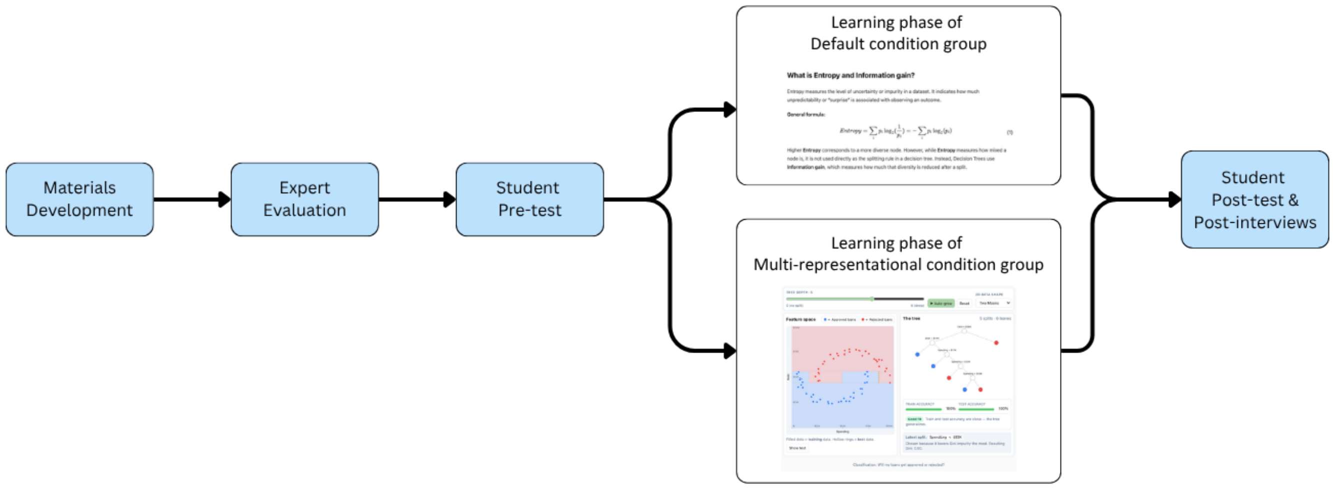


Figure 1: Overview of this research's methodology.

3.3 Instructional Material Design

Two instructional setups were developed using Jupyter Book, both aligned with identical learning objectives and topic coverage but differing in the combination of instructional representations employed. The design principles used for these teaching materials were based on: Cognitive Load Theory [6], Multimedia Learning Theory [5], and constructivist learning principles.

The instructional materials for this tutorial were formed using a structured instructional design approach combining the above derived learning objectives from Bloom's Revised Taxonomy, and the **ADDIE framework** [18]:

- **Analysis**: Core Decision Tree concepts were identified, including tree structure, recursive partitioning, and node impurity-based splitting.
- **Design**: Learning objectives were defined with Bloom's Revised Taxonomy to structure cognitive progression.
- **Development**: Instructional content was created using step-by-step explanations, visual intuition, and guided examples of splitting and impurity computation.
- **Implementation**: The tutorial was implemented as a self-contained learning resource.
- **Evaluation**: The final Jupyter notebooks have been revised after being reviewed by a few Master's students and Teaching Assistants from the EEMCS department.

Instructional materials for Default Condition

The material for this setup consists of mainly textual explanations, static examples, and mathematical formulas without the inclusion of interactive or multimedia elements.

Instructional materials for Multi-representational Condition

The material for this setup includes: Visual Decision Tree diagrams, Interactive simulations, Animated split demonstrations, Analogies for different splitting criteria concepts (Entropy & Information gain, Gini impurity), and Video explanations.

3.4 Experimental Design

In order to test the initial hypotheses and answer the Research Question, a pilot between-subject experiment design was employed, measuring participants' conceptual understanding, problem-solving performance, and knowledge transfer on Decision Trees.

This study was designed as a pilot study to explore the potential effects of the intervention and assess the feasibility of the experimental setup. Due to the small sample size (N=10), divided across two groups, the study was not powered for robust statistical inference or hypothesis testing. Instead, the analysis focused on identifying preliminary trends and generating insights to inform future large-scale studies.

Participants. This research study recruited 10 students of the first-year Computer Science and Engineering (CSE) Bachelor's programme at Delft university of Technology. They are specifically chosen because, as first-year students, they are the most likely to not have followed any ML courses in the CSE programme before. Nevertheless, they were still asked during recruitment if they have any prior experience in ML, this is to ensure that students that already have exposure to ML concepts are excluded from the experiment.

Procedure. Participants were randomly assigned to the two conditions using variable block randomization to keep group sizes balanced while remaining unpredictable during recruitment: the Default condition group (N=5), and the Multi-representational condition group (N=5). The procedure is then conducted within a duration of around 60 minutes, in the following order:

- **Quantitative pre-test assessment**: This ensures learners all have the same baseline level of mathematical foundations and exclude those with prior Decision Trees knowledge, all participants have completed a pre-test. The pre-test assessment consisted of three sections:

Probability Theory, Logarithms, and Decision Tree concepts. The first two sections were designed to assess students' understanding of the mathematical foundations required for learning Decision Trees. The final section served as a screening tool to identify students who had not been previously exposed to Decision Tree concepts. Participants were allowed to use a calculator, **except for the Logarithms section**, this is to accurately assess their knowledge in Logarithms. The questions for this pre-test assessment are included in Appendix A

- **Learning phase**: All participants carry out a structured learning sessions using the mentioned above tutorials, according to their assigned group, carried out in a quiet space without distractions to provide the best study environment for participants.
- **Quantitative post-test assessment**: The post-test measured learners' conceptual understanding, problem solving performance, and knowledge transfer about Decision Trees. Conceptual understanding was assessed through questions on Decision Tree structures, Supervised Learning models, Gini impurity, Entropy, Information gain, tree construction, and Overfitting vs Underfitting. Problem-solving performance was evaluated through tasks involving manual Decision Tree construction using different splitting criteria and dataset-based decision making. Knowledge transfer was assessed by requiring participants to apply Decision Tree concepts to previously unseen data.The questions for this pre-test assessment are included in Appendix B
- **Semi-structured interview**: Qualitative insights about participants' reflections on their overall thought process, and feedbacks on the tutorial were gathered through semi-structured interviews. The questions for this pre-test assessment are included in Appendix C

Measurements. A pre-test is used to measure the the baseline mathematical and logical reasoning ability, including probability, logarithms, and basic tree structures. After that, to measure whether there are any difference between the Default Condition group and the Multi-representational Condition group, a post-test was administered, with three sections for Conceptual understanding, Problem-solving performance, and Knowledge transfer. Assessment for both pre-test and post-test items include multiple-choice and short-answer questions scored using a rubric:

- 0 = incorrect or no understanding
- 1 = partial understanding
- 2 = mostly correct but incomplete explanation
- 3 = fully correct and well-reasoned explanation

The final score is normalized to a 0–100 scale:

$$\text{Final score} = \frac{\text{Raw Score}}{\text{Max Score}} \times 100 \tag{1}$$

In addition, the pre-test and post-test both include the same set of questions for Confidence Ratings (0-10) about Decision Trees core concepts, this is to review whether completing the tutorial enhances students' level of confidence in the topics.

Followed the post-test, each participant went through a semi-structured interview, where they were asked the overall learning experience using their assigned tutorial, and were asked to describe the core concepts in Decision Trees that they have just learned, along with their though process.

Data analysis. This study evaluates learning outcomes in Decision Tree education through three primary dependent variables: conceptual understanding, problem-solving performance, and knowledge transfer ability. Baseline mathematical reasoning, measured via a pre-test, all variables are assessed using rubric-based scoring and analyzed through descriptive statistics (mean, median, standard deviation, minimum, maximum) and inferential tests, including t-tests and, where t-test results are non-significant, Mann-Whitney U tests. This ensures clear summary of the data and reliable comparison between groups using suitable statistical tests.

Semi-structured interviews were analyzed using both inductive and deductive thematic analysis to examine participants' learning experiences across the two instructional conditions. The inductive analysis involved open coding of participant responses followed by iterative grouping into higher-order themes. Deductive analysis subsequently mapped these themes onto established theoretical frameworks including Cognitive Load Theory, Dual Coding Theory, and Multimedia Learning Theory. Qualitative data were analyzed using a hybrid approach combining deductive and inductive thematic analysis, following Braun and Clarke's six-phase framework [19]: (1) familiarization with data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report.

Deductive Analysis. A deductive coding framework was developed prior to analysis, derived from the study's theoretical foundations (Cognitive Load Theory, Dual Coding Theory, Multimedia Learning Theory) and research questions. The framework comprised six a prior themes with explicit inclusion criteria (see Appendix D. for the complete coding framework). Transcripts were coded against these themes by the primary researcher. This approach is appropriate for this study because it helps check whether the data fits existing learning theories in a structured and consistent way.

Inductive Analysis. The inductive analysis revealed four emergent themes that captured participants' learning experiences beyond the theoretically-derived frame- work, providing subtle insights into how learners engaged with the instructional materials across conditions. Following deductive coding, transcripts were subjected to open inductive coding to identify emergent themes not captured by the theoretical framework. This involved line-by-line coding of participant responses, followed by iterative grouping of codes into higher-order themes (see Appendix E for the complete inductive coding structure). This method helps discover new patterns in the data that were not expected from existing theories.

4 Results

Given the relatively small sample size of this study (N=10), the findings should be interpreted as preliminary and exploratory rather than conclusive. As such, this research functions as a pilot experiment, intended to identify potential

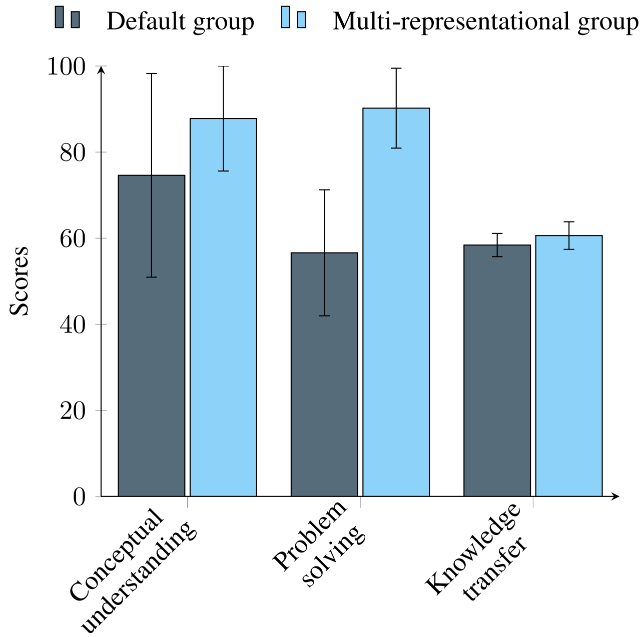


Figure 2: Average scores of the post-test between 2 groups.

trends, patterns, and relationships within the data.

While the current sample limits statistical power and generalizability, the observed results provide an initial indication of possible effects that warrant further investigation. With a larger sample size, future research could apply more robust statistical analyses, increase the reliability and precision of estimates, and better assess whether the trends observed in this pilot study persist across a broader population.

4.1 Baseline Knowledge in pre-test

A total of 10 participants completed the study, with 5 participants assigned to each instructional condition. Baseline mathematical and logical reasoning abilities were assessed using the pre-test described in Section 3.3.

Group	Mean	SD	Min	Max
Default Instruction	72.2	6.94	61	78
Multi-representational	74.0	9.57	60	83

Table 2: Pre-test descriptive statistics

Both t-test ($p\_value = 0.74$) and Mann–Whitney U test ($U = 11.0$, $p = 0.69$) showed **no statistically significant difference** between groups on pre-test scores, indicating comparable baseline knowledge.

4.2 Confidence levels

Participants showed an overall increase in mean confidence from pre-test ($M \approx 0.70/10$) to post-test ($M \approx 3.12/10$). The largest gains were observed in Decision Tree concepts, with Entropy/Information gain increasing from $M \approx 0.5$ to $M \approx 4.2$ and Gini impurity from $M \approx 0.5$ to $M \approx 5.8$. General ML and Supervised Learning confidence also improved modestly, from around $M \approx 2.2$ to $M \approx 3.0$ across items. Although both groups improved, **the multi-representation group showed slightly larger mean increases overall, but results remain exploratory due to the small sample size**.

4.3 Conceptual Understanding

Participants who received multi-representational instruction achieved higher conceptual understanding scores than participants in the default condition.

Group	Mean	Median	SD
Default Instruction	74.6	62	23.66
Multi-representational	87.8	88	12.20

Table 3: Conceptual understanding scores

Both t-test ($p\_value = 0.30$) and Mann–Whitney U test ($U = 8.0$, $p = 0.31$) showed **no statistically significant difference** between groups on conceptual understanding scores, indicating there are no significant difference between their level of conceptual understanding in Decision Trees.

4.4 Problem-Solving Performance

Problem-solving performance was evaluated through tasks involving split selection, impurity calculations, and manual tree construction.

Group	Mean	Median	SD
Default Instruction	56.6	63	14.64
Multi-representational	90.2	87	9.28

Table 4: Problem-solving performance scores

The difference between groups was statistically significant according to a t-test ($p\_value = 0.0025$) and a Mann–Whitney U test ($U = 0.0$, $p = 0.0079$).

Participants exposed to multiple representations witnessed a much higher score in applying Decision Tree concepts to structured problem-solving tasks.

4.5 Knowledge Transfer

Knowledge transfer was measured using novel classification scenarios that differed from those encountered during instruction.

Group	Mean	Median	SD
Default Instruction	58.4	58	2.70
Multi-representational	60.6	61	3.21

Table 5: Knowledge transfer scores

Both t-test ($p\_value = 0.27$) and Mann–Whitney U test ($U = 7.0$, $p = 0.22$) showed **no statistically significant difference** between groups on knowledge transfer scores, indicating there are no significant difference between their level of knowledge transfer in Decision Trees.

4.6 Qualitative Findings

Semi-structured interviews were conducted immediately following the post-test, with each participant completing a 5-minute interview. Questions centered on participants' learning experiences, their understanding of Decision Tree concepts, and their feedback on the instructional materials. Participants were encouraged to elaborate on any additional topics they deemed relevant.

Group A: Default condition

Group B: Multi-representational condition

Deductive thematic analysis. The most frequently observed theme was Conceptual Visualization and Understanding (9/10 participants), followed by Knowledge Transfer and Application (10/10 participants), indicating that both groups engaged substantially with visual and applied aspects of the learning materials. However, important group differences emerged:

- Cognitive Load and Mental Effort: Participants in Group B reported higher instances of cognitive strain (3/5 participants) compared to Group A (2/5 participants), suggesting that exposure to multiple representations may have increased processing demands, despite potentially enhancing understanding. One participant from group B reported that the video meant to explain the Entropy concept was "was less intuitive than expected", and another participant from group B reported that the whole learning tutorial was "harder than [*they*] expected". While the two participants from group A expressed that they expected to overall the learning tutorial would take less time.
- Instructional Design Quality: Group B participants (4/5 participants) provided more feedback on instructional design features than Group A (2/5 participants), indicating greater engagement with the multi-representational approach and its varied components. One interesting comment from a group B participant was they found it a bit difficult to connect theories to application in the beginning, but after completing the tutorial they could see how everything connected to each other and found that the visualizations and analogies really helped them connecting the dots.

Inductive thematic analysis. The inductive analysis revealed four emergent themes that captured participants' learning experiences beyond the theoretically-derived framework, providing subtle insights into how learners engaged with the instructional materials across conditions.

Multi-representations Enhanced Understanding (8/10 participants) emerged as the most prevalent theme, with a striking group difference: all Group B participants (5/5) reported that multi-representational features supported their understanding. As one Group B participant noted, "the interactive simulation part showed different data shapes and how they were split really helped their understanding". Another remark from a participant in Group B was "The diagram for different types of Machine Learning made it easier for me to understand". This theme, combining visual/interactive support and helpful analogies, indicates that the multi-representational condition may provide distinctive pedagogical value.

5 Responsible Research

5.1 Ethical concerns

This study involves human participants in an educational intervention context, and therefore ethical considerations and methodological transparency are essential. The research was designed to minimize potential risks, ensure voluntary participation, and support reproducibility of the experimental procedure.

Participation in this study was fully voluntary, and individuals were informed about the purpose of the research prior to participation. Participants were made aware that they could withdraw at any time without penalty. No personal identifying information was collected, responses were anonymized during analysis to ensure privacy and confidentiality.

The study does not involve sensitive personal data, medical information, or high-risk interventions. However, even in low-risk educational research, ethical responsibility is required to prevent indirect harm such as cognitive overload, frustration, or perceived academic evaluation pressure.

To mitigate these risks:

- Participants were not graded in a way that affected academic standing.
- Instructions emphasized that the study was focused on learning processes rather than individual performance judgment.

5.2 Reproducibility

To support replication and verification of this pilot study, all instructional materials (including the Jupyter Book tutorials for both conditions, assessment instruments (pre-test, post-test, interview protocol), anonymized participant data, and analysis code) are made available. The two Jupyter notebooks and codes for codes to run statistical tests are released openly. As for the other experiment design such as pre-test, post-test, and interview questions, as well as the thematic coding framework for qualitative analysis; these are all provided within the appendices of this paper.

The experimental procedure is documented step-by-step, including recruitment criteria, randomization method, environmental conditions, and scoring rubrics. Deductive and inductive coding frameworks for qualitative analysis are fully specified in the appendices. However, the pilot study status may require procedural refinements for larger-scale replication. Future studies should prioritize larger samples, longitudinal designs, and multi-context replications to validate and extend these preliminary findings. In addition, the coding for thematic data analysis was conducted solely by the primary researcher, future replications and reanalysis should consider having multiple coders to cross-examine their methods and produce more reliable and credible qualitative results.

6 Discussion

This study investigated whether multi-representational instructional materials improve learning outcomes in Decision Tree education compared with traditional text-based materials. Across all three measured outcomes (conceptual understanding, problem-solving performance, and knowledge transfer) participants in the multi-representational condition achieved slightly higher scores than those in the default condition.

The strongest improvement was observed in problem solving performance. The significant improvement in problem-solving performance aligns with qualitative reports from Group B participants, who specifically highlighted the interactive simulations as helpful for understanding how data splits work (Theme 1 in inductive analysis). This suggests that the interactive components may have been particularly effective for procedural learning and application tasks.

The finding that Group B participants reported higher cognitive strain (3/5 vs 2/5) while also achieving better problem-solving performance suggests a trade-off: additional representations may increase mental effort but also enhance understanding of procedural concepts. This aligns with Sweller's concept of 'germane cognitive load', effort that contributes to learning rather than extraneous load that hinders it [20].

The findings provide partial support for the research hypotheses. While problem-solving performance showed a statistically significant improvement ($p = 0.0025$), differences in conceptual understanding ($p = 0.30$) and knowledge transfer ($p = 0.27$) did not reach statistical significance, though trends favored the multi-representational condition. **However, as this was only a pilot study with a small sample size, the results should be interpreted cautiously.** Rather than providing definitive conclusions, the study primarily serves to explore feasibility, identify preliminary patterns, and inform the design of larger, more rigorous future experiments.

Limitations. Several limitations should be considered when interpreting these findings. First, the sample size was very small ($N = 10$), limiting statistical power and reducing the generalisability of the results. Second, the intervention consisted of a single learning session, making it impossible to assess long-term retention. Third, individual differences in motivation, prior machine learning exposure, and learning preferences may have influenced performance. Furthermore, the thematic analysis was coded exclusively by the primary researcher; therefore, future studies should involve multiple coders to compare approaches and enhance the reliability and credibility of the qualitative findings. Finally, the study focused exclusively on Decision Trees and therefore cannot establish whether similar effects would be observed for more advanced machine learning models.

Despite these limitations, the study provides preliminary evidence that carefully designed multi-representational instructional materials can improve introductory ML education. The findings also demonstrate the feasibility of combining quantitative and qualitative methods to evaluate educational interventions in machine learning.

7 Conclusions and Future Work

This study investigated the extent to which multi-representational instruction improves conceptual understanding, problem-solving performance, and knowledge transfer in Decision Tree learning compared to traditional text-based materials.

Using a mixed-methods experimental design, participants learned Decision Tree concepts through either a text-only tutorial or a tutorial incorporating visualizations, analogies, videos, and interactive simulations. Learning outcomes were evaluated using measures of conceptual understanding, problem-solving performance, and knowledge transfer.

The results suggest that multi-representational instruction may provide meaningful benefits across all measured outcomes. Participants exposed to multiple representations achieved comparable conceptual understanding, knowledge transfer scores. However, they performed better on Decision Tree problem-solving tasks to unfamiliar situations. Due to the small sample size of this research, these findings should not be treated as definitive conclusions but rather this study's main purpose is to explore feasibility, potential trends to inform the design of a larger future experiments

The primary contribution of this study is the design and evaluation of a structured multi-representational learning environment for Decision Tree education. In addition, the study proposes a pilot experiment of mixed-method evaluation framework that combines quantitative outcome measures with qualitative learner experiences, providing a foundation for future research in machine learning education.

Future work should replicate the study using larger participant samples to improve statistical power and external validity. Longitudinal studies could investigate whether the observed benefits persist over time and influence long-term retention. Further research could also explore different combinations of representations and examine whether similar effects occur when teaching more advanced machine learning topics, such as Random Forests, Support Vector Machines, and Neural Networks.

Overall, the findings provide preliminary evidence that well-designed multi-representational instruction may have the potential to improve the teaching and learning of foundational machine learning concepts and may serve as a promising direction for future curriculum development in machine learning education.

References

[1] Victor F.C. Souza, Ferdinando Cicalese, Eduardo Sany Laber, and Marco Molinaro. Decision trees with short explainable rules. *Theoretical Computer Science*, 1047:115344, 2025.

[2] Tim Erickson and Joachim Engel. What goes before the cart? introducing classification trees with arbor and codap. *Teaching Statistics*, 45(S1):S104–S113, 2023.

[3] Yannik Fleischer and Rolf Biehler. Exploring students' constructions of data-based decision trees after an introductory teaching unit on machine learning. *ZDM – Mathematics Education*, 04 2025.

[4] Susanne Podworny, Rolf Biehler, and Yannik Fleischer. Young students' engagement with data to create decision trees. *ZDM*, 01 2025.

[5] Richard E. Mayer. *Multimedia Learning*. Cambridge University Press, 2nd edition, 2009.

[6] Jan L Plass, Roxana Moreno, and Roland Brünken. Cognitive load theory. 2010.

[7] R. K. Atkinson. Multimedia learning of mathematics. In R. E. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, pages 393–408. Cambridge University Press, 2005.

[8] R. Tarmizi and J. Sweller. Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80(4):424–436, 1988.

[9] R. M. Martins and C. Gresse von Wangenheim. Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education*, 22(3):421–446, 2023.

[10] M. H. Kaspersen, K. E. K. Bilstrup, M. Van Mechelen, A. Hjort, N. O. Bouvin, and M. G. Petersen. The machine learning machine: A tangible user interface for teaching machine learning. In *Proceedings of the 20th ACM Interaction Design and Children Conference*, pages 1–13. ACM, 2021.

[11] A. Lichtenberger, T. Kokkonen, and L. Schalk. Learning with multiple external representations in physics: Concreteness fading versus simultaneous presentation. *Journal of Research in Science Teaching*, 61(9):2258–2290, 2024.

[12] T. Kokkonen, A. Lichtenberger, and L. Schalk. Concreteness fading in learning secondary school physics concepts. *Learning and Instruction*, 77:101524, 2022.

[13] O. Hazzan and K. Mike. *Guide to Teaching Data Science: An Interdisciplinary Approach*. Springer, 2023.

[14] L. S. Marques, C. Gresse von Wangenheim, and J. C. R. Hauck. Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education*, 19(2):283–321, 2020.

[15] S. D. Essinger and G. L. Rosen. Machine learning: An introductory unit of study for secondary education. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, pages 1–7. ACM, 2019.

[16] X. Wan, X. Zhou, and J. M. Zydney. Smileycluster: Supporting accessible machine learning in k-12 scientific discovery. In *Proceedings of the Interaction Design and Children Conference*, pages 1–12. ACM, 2020.

[17] Mary Forehand. Bloom's taxonomy. *Emerging perspectives on learning, teaching, and technology*, 41(4):47–56, 2010.

[18] Robert Maribe Branch and İlhan Varank. *Instructional design: The ADDIE approach*, volume 722. Springer, 2009.

[19] Victoria Clarke and Virginia Braun. Thematic analysis. *The journal of positive psychology*, 12(3):297–298, 2017.

[20] John Sweller. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, 22(2):123–138, 2010.

Appendices

The appendices include questions from the pre-test, post-test, and interviews; as well as deductive and inductive coding structures for qualitative thematic data analysis.

A Pre-test questions

A.1 Probability Theory

1. A fair coin is flipped 5 times. What is the probability of getting exactly 3 heads?
   (A) $\frac{10}{16}$
   (B) $\frac{5}{32}$
   (C) $\mathbf{\frac{10}{32}}$
   (D) $\frac{5}{12}$
   (E) I don't know
2. A fair coin is flipped 5 times. What is the probability of getting exactly 1 tail?
   (A) $\mathbf{\frac{5}{32}}$
   (B) $\frac{1}{16}$
   (C) $\frac{15}{16}$
   (D) $\frac{31}{32}$
   (E) I don't know
3. A box contains 7 apples, 2 cherries, and 1 orange. You randomly pick up a fruit from the box. What is the probability of getting a cherry?
   (A) $\frac{1}{2}$
   (B) $\mathbf{\frac{1}{5}}$
   (C) $\frac{1}{10}$
   (D) 1
   (E) I don't know
4. There are 2 boxes containing the same number of fruits.
   Box 1: 7 apples, 2 cherries, and 1 orange.
   Box 2: 7 apples, 2 cherries, and 1 orange.
   You randomly pick up 2 fruits, 1 from each box. What is the probability that the 2 fruits you picked up are the same type?
   (A) $\frac{14}{23}$
   (B) $\mathbf{\frac{27}{50}}$
   (C) $\frac{49}{250}$
   (D) 1
   (E) I don't know
5. A medical test is given to patients. The test correctly identifies sick patients 95% of the time and correctly identifies healthy patients 90% of the time. If you pick a patient at random and perform the test: Are the events "the test result is positive" and "the patient actually has the disease" independent events?
   (A) Yes
   (B) **No**
   (C) It depends
   (D) I don't know

6. In an email system:
40% of emails are spam.
80% of spam emails contain the word "FREE".
10% of non-spam emails contain the word "FREE".
If an email is randomly selected and it is known to contain the word "FREE", what is the probability that it is spam?
   (A) 0.32
   (B) $\mathbf{0.842}$
   (C) 0.38
   (D) 0.921
   (E) I don't know

A.2 Logarithms

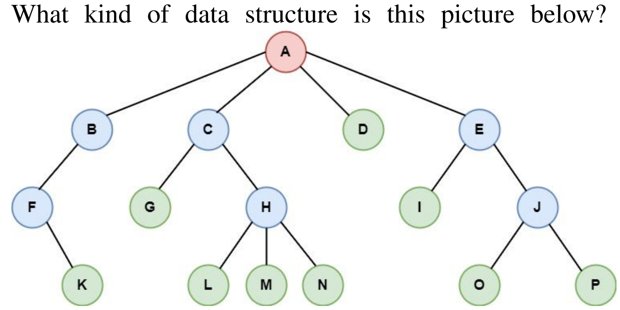
7. Evaluate this expression: $\log_2(32) = ?$
   (A) 3
   (B) 2
   (C) $\mathbf{5}$
   (D) 4
   (E) I don't know
8. Evaluate this expression: $\log_{10}(1000) = ?$
   (A) 10
   (B) 5
   (C) 2
   (D) $\mathbf{3}$
   (E) I don't know
9. Solve for x: $2^x = 64$
   (A) x = 64
   (B) x = 7
   (C) x = 32
   (D) **x = 6**
   (E) I don't know
10. Evaluate this expression. $log_2(0.5) = ?$
Answer: -1
11. Which value of p would maximize $f(p)$? Explain your answer.
$f(p) = -[p\log_2(p) + (1-p)\log_2(1-p)]$
12. Compute this equation below.
$-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$
Answer: 1
13. Select all that applies.
    - $\log_a(X \times Y) = \log_a(X) \times \log_a(Y)$
    - $\mathbf{\log_a\left(\frac{M}{N}\right) = \log_a(M) - \log_a(N)}$
    - $\log_a(1) = 1$
    - $\log_a(X^k) = a \times \log_k(X)$
    - $\mathbf{\log_a(1) = 0}$
    - $\mathbf{\log_a(X \times Y) = \log_a(X) + \log_a(Y)}$
    - $\log_a\left(\frac{M}{N}\right) = \frac{\log_a(M)}{\log_a(N)}$
    - $\log_a(a) = a$
    - $\mathbf{b^{\log_b(k)} = k}$
    - $\mathbf{\log_a(X^k) = k \times \log_a(X)}$
    - $\mathbf{\log_a(a^k) = k}$

A.3 Confidence Ratings (Slider 0–10)

14. How confident are you with topics in Machine Learning?
15. How confident are you with topics in Supervised Learning models?
16. How confident are you with Classification tasks?
17. How confident are you with topics in Decision Trees?
18. How confident are you with Entropy and Information gain?
19. How confident are you with Gini impurity?

A.4 Basic Tree structures

20. If you are somewhat confident or extremely confident with any topics above, please provide shortly what you know below. Otherwise, you can skip this question.
21. What kind of data structure is this picture below?



22. What kind of node is node M?
    (A) Root node
    (B) Internal node
    (C) **Leaf node**
    (D) M is not a node
    (E) I don't know
23. What kind of node is node A?
    (A) **Root node**
    (B) **Internal node**
    (C) Leaf node
    (D) A is not a node
    (E) I don't know
24. What kind of node is node J?
    (A) Root node
    (B) **Internal node**
    (C) Leaf node
    (D) J is not a node
    (E) I don't know

B Post-test questions

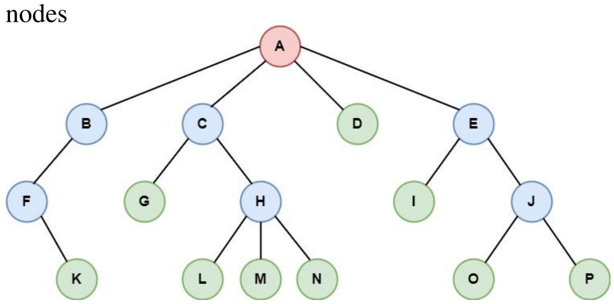
B.1 Confidence Ratings (Slider 0–10)

1. How confident are you with topics in Machine Learning?
2. How confident are you with topics in Supervised Learning models?
3. How confident are you with Classification tasks?
4. How confident are you with topics in Decision Trees?

5. How confident are you with Entropy and Information gain?
6. How confident are you with Gini impurity?

B.2 Conceptual understanding

7. Please list out all root nodes, internal nodes, and leaf nodes



8. You are computing the Gini impurity as a splitting criteria of a continuous (numerical) feature: Age of a person. What is the first thing you should do?
   (A) Calculate Entropy for each age
   (B) Calculate Gini impurity for each age
   (C) **Sort the data in ascending order**
   (D) Calculate the average age to split the data into 2
   (E) **Sort the data in descending order**
9. What does Gini impurity measure?
   (A) The accuracy of a model
   (B) How many times a Decision Tree split
   (C) **The randomness of a node**
   (D) The number of features in the dataset
10. Explain in your own words the intuition behind the Gini impurity formula.
11. Explain in your own words the intuition behind the Entropy formula.
12. When is a node considered to be pure?
   (A) It has maximum entropy
   (B) It contains many features
   (C) **It contains only 1 feature**
   (D) It has an equal number of all classes
13. Which statement about Entropy is correct?
   (A) Entropy is lowest when classes are evenly distributed
   (B) **Entropy measures class uncertainty in a node**
   (C) Entropy is highest when all samples belong to one class
   (D) Entropy measures the reduction in impurity after a split

B.3 Problem-solving

14. In a house, there lives 2 humans, 1 dog, and 1 cat. Calculate the Entropy of this house, give your answer to 3 decimal places.
    Answer: 1.500
15. A node contains 8 data points from class A, and 2 data points from class B. Calculate the Entropy of this node, give your answer to 3 decimal places. Is this node relatively pure or impure?
    Answer: 0.722, relatively pure
16. A parent node has entropy = 1.0. Two possible splits produce:
    - Split A: Left child entropy = 0.2, Right child entropy = 0.3
    - Split B: Left child entropy = 0.7, Right child entropy = 0.8

    Which split is better? And which split yields higher Information gain?
    Answer: Split A is better because it produces much "purer" children (lower entropy), so it reduces uncertainty far more than Split B.

B.4 Knowledge transfer

17. A Decision Tree that continues splitting until every leaf contains one sample is most likely:
    (A) Underfitting due to too much simplification
    (B) **Overfitting due to memorization of training data**
    (C) Optimal because it eliminates all bias
    (D) Robust to unseen data because leaves are pure
18. You are building a decision tree to predict whether a customer will buy a product ("Yes" or "No"). After the first split, one group contains 99 buyers and 1 non-buyer, while the other group contains 50 buyers and 50 non-buyers. Which statement best describes the split?
    (A) The split is bad because both groups still contain some non-buyers.
    (B) **The split is good because the first group is very pure, even though the second group is not.**
    (C) The split is useless because one group is much larger than the other.
    (D) The split is perfect because it separated all buyers from non-buyers.
19. A car insurance company builds a decision tree to predict which drivers will file a claim. The first split is on "Age ≤ 25?" and the second split (for drivers over 25) is on "Annual mileage $> 30,000$?". Why might the tree NOT split on "Annual mileage" for drivers under 25?
    (A) The tree forgot to consider that feature for that group.
    (B) **Among drivers under 25, mileage did not help separate claim-filers from non-filers.**
    (C) The "Annual mileage" feature only applies to drivers over 25.
    (D) The tree always uses the same feature for all branches.

C Interview questions

1. How was the overall learning experience?

2. Was there any part in the tutorial that you skipped, or you feel like you could have skipped?
3. Can you explain a Decision Tree to someone who has never heard of this concept before?
4. Can you explain Gini impurity in your own words?
5. Can you explain Entropy in your own words?
6. Do you have any other feedback regarding this research study?

D Deductive thematic qualitative codes

Qualitative post-learning data gathered through semi-structured interviews

(with identified **deductive** codes)

Group A: Static/Default condition

Group B: Multi-representational condition

Based on the thesis literature review and theoretical foundations, the following deductive coding framework organized around six key themes derived from prior research and the study's research questions:

No.	Themes	Definition	Sources	Prevalence
1	Cognitive Load and Mental Effort	This captures participants' references to mental effort required to process instructional materials, including statements about difficulty, time investment, cognitive strain, working memory demands	Cognitive Load Theory [6]	5/10 **Group A**: 2 **Group B**: 3
2	Conceptual Visualization and Understanding	This captures participants' references to visuals, analogies, or interactive elements that supported conceptual understanding	Cognitive Theory of Multimedia Learning [5]	9/10 **Group A**: 4 **Group B**: 5
3	Engagement and Motivation	This captures participants' references to their level of interest, motivation, and engagement with the learning materials	Dual Coding Theory [7]	5/10 **Group A**: 3 **Group B**: 2
4	Level of Conceptual Understanding	This captures participants' references to ability to apply, transfer, and explain Decision Tree concepts in novel contexts and their own words	Research hypotheses H_2 and H_3	10/10 **Group A**: 5 **Group B**: 5
5	Instructional Design Quality	This captures participants' references to specific instructional design features, both positive and negative	ADDIE framework [18]	6/10 **Group A**: 2 **Group B**: 4
6	Conceptual Difficulties	This captures participants' references to explicit identification of conceptual difficulties, struggles, or misunderstandings	Fleischer & Biehler [3], Podworny et al. [4]	3/10 **Group A**: 1 **Group B**: 2

E Inductive thematic qualitative codes

Qualitative post-learning data gathered through semi-structured interviews

(with identified **inductive** codes)

Group A: Static/Default condition

Group B: Multi-representational condition

The following initial candidate themes were identified through inductive coding following Braun and Clarke's (2006) six-phase framework:

No.	Themes	Indication
1	Visual and Interactive Support	→ Helpfulness of diagrams, simulations, and visualizations
2	Helpful Analogy and Worked example approach	→ Helpfulness of analogies, and worked examples
3	Conceptual Gaps	→ Missing explanations, unclear concepts
4	Expectation Mismatch	→ Difficulty level vs. expectations, time investment
5	Knowledge Integration	→ Connecting theory to practice, mapping to model behavior
6	Learning Preferences	→ Desire for more examples, interactivity, real-world applications
7	Metacognitive Awareness	→ Self-report of skipped content, skimming behavior
8	Correctness in Concepts Definition	→ Explain correctly the some basic concepts in Decision Tree

The candidate theme 7 was omitted due to its irrelevancy to this research, as all content that was either skipped or skimmed belonged to the first section Mathematical foundations revising Probability and Logarithms for students. This chapter only serves as a refresher as needed for students, and not part of the Research Question.

After review and evaluation, the candidate themes were grouped into 4 main important themes as follows:

Themes	Combined candidate themes	Prevalance	Group A (N=5)	Group B (N=5)
Multi-representations enhanced understanding	1, 2	8/10	3/5	5/5
Bridging knowledge gaps	3, 5, 6	4 /10	2/5	2/5
Expectation vs Cognitive mismatch	4	4 /10	2/5	2/5
Basic Conceptual Understanding	8	10/10	5/5	5/5