Teaching Principal Component Analysis Through Multiple Representations

Impacts on Conceptual Understanding, Problem Solving and Knowledge Transfer

Iulia-Cristiana Vlădăreanu[1]

Supervisors: 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: Iulia-Cristiana Vlădăreanu
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Ilinca Rențea, Jorge Martinez Castaneda

Abstract

Machine Learning has become a prominent component of STEM curricula, yet there remains a lack of standardized pedagogical frameworks for teaching it. Principal Component Analysis represents a typical educational challenge, requiring students to simultaneously coordinate algebraic manipulation, geometric intuition, and algorithmic thinking. This study evaluates how different combinations of instructional representations affect undergraduate students' conceptual understanding, problem-solving performance, and knowledge transfer ability. An experiment was conducted with 27 first-year Computer Science students at TU Delft. Participants were assigned to one of two experimental conditions: a traditional static representations group or a multimedia-enhanced representations group. Quantitative analysis revealed that the static group significantly outperformed the interactive group in total post-test scores and knowledge transfer. No statistically significant differences were observed in conceptual understanding or problem-solving performance. Qualitative thematic analysis indicated a disconnect between perceived and actual learning: while students preferred the interactive widgets for building geometric intuition, these features may have introduced extraneous cognitive load, provided a false sense of understanding, or students simply ran out of time because interactive exploration takes longer.

1 Introduction

In recent years, the field of Machine Learning (ML) has been rapidly integrated into both industrial applications - such as predictive maintenance [11], quality inspection [19], or industrial sensing [10] - and academic research, with applications in domains such as medicine [33], astrophysics [30], or materials science [21]. Consequently, ML has become a prominent component of undergraduate and graduate STEM curricula. As it becomes a foundational competence for engineers and scientists, the quality of ML education is proving increasingly crucial. However, this rapid growth has not yet been matched by a corresponding body of empirical educational research. The academic community currently lacks standardized pedagogical methods and comprehensive frameworks for ML instruction, leaving educators with limited research-based support for curriculum and instructional design.

One of the central challenges in ML education lies in the abstract mathematical foundations required to understand core ML concepts and techniques. Topics such as backpropagation, regularization, and dimensionality reduction often rely heavily on knowledge from multivariable calculus, linear algebra and high-dimensional geometry. As a result, students are required to coordinate algebraic manipulation, geometric intuition, and algorithmic thinking simultaneously, which can become a barrier for those lacking strong mathematical skills. [28]

Principal Component Analysis (PCA) provides a representative case of such a concept. It is a widely used dimensionality reduction technique that reduces data complexity while preserving the most important features in a dataset. PCA is applied across a broad range of domains for purposes such as data compression, visualization, noise reduction and feature extraction.

PCA is a particularly interesting topic for educational research because understanding it requires concepts from multiple areas of mathematics, including linear algebra, statistics, and geometry. Existing learning materials therefore combine concepts such as covariance matrices, eigenvectors, eigenvalue decomposition, variance, and vector projection to explain the method [27; 6]. At the same time, PCA can also be introduced through intuitive analogies and concrete examples [23], making the underlying ideas accessible to students with limited mathematical backgrounds. Moreover, the geometric interpretation of PCA allows concepts such as dimensionality reduction, projection, and variance maximization to be represented visually through diagrams, animations, and interactive widgets.

More broadly, research in multimedia learning suggests that understanding can be enhanced when concepts are presented through coordinated representations. Atkinson [2] emphasizes the role of multimedia learning in mathematics education, while Lichtenberger et al. [12] demonstrate the impact of multiple representations in physics. While similar findings exist in other subjects such as chemistry, biology, and even language acquisition [8; 7; 17], their implications for ML education remain mostly unexplored, leading to a clear gap in the literature.

The objective of this research paper is to evaluate how different combinations of instructional representations affect undergraduate students' understanding of PCA. By using PCA as a case study, this study addresses the following research question:

"How does the combination of multiple instructional representations across the full learning pipeline - from prerequisite mathematical concepts to PCA instruction - affect students' conceptual understanding, problem-solving performance, and knowledge transfer ability when learning PCA?"

This research directly builds on the recent work of Rențea et al. [20], who explored the effect of interactive visualizations on teaching ML topics to undergraduate Computer Science and Engineering students. Their comparison of static versus interactive groups produced mixed results; while interactive visualizations led to a significant increase in knowledge gain for PCA, they found no significant differences in knowledge gain for Gradient Descent, nor did they observe any significant effects on student motivation across either topic. This study directly follows up on their findings for PCA by contributing additional data and a comprehensive mixed-methods design. Specifically, the main contributions of this research are:

- **Comprehensive evaluation of learning outcomes**: Moving beyond knowledge gain, this study evaluates the impact of instructional representations across three distinct cognitive processes - conceptual understanding, problem solving, and knowledge transfer.

- **Connected qualitative and quantitative insights**: Through a mixed-methods design, this study aims to uncover relations between perceived and actual student learning, as well as cognitive load effects of these multiple representations.
- **Prerequisite focus and full-pipeline instruction**: This study examines the combination of multiple instructional representations across a complete learning pipeline, starting with a review of the necessary mathematical prerequisites and continuing with a complete tutorial of PCA.

The structure of this paper is as follows: Chapter 2 (Related Works) explores the current literature on ML education, mathematical challenges students face and the integration of interactive learning in instructional materials. Chapter 3 (Methodology) outlines the conceptual background of PCA, the development of the instructional materials used in the experiment and the experimental design. Chapter 4 (Results) presents both the quantitative and qualitative findings from the experiment. Chapter 5 (Discussion) interprets these results and examines both student preferences and objective performance. Chapter 6 (Responsible Research) expands upon ethical considerations, data privacy and experiment reproducibility. Chapter 7 (Conclusion) summarizes the core findings of the research and proposes directions for future studies.

2 Related Works

To position this work within the existing literature, relevant prior research is reviewed below.

2.1 Pedagogy and Frameworks in ML Education

Although Machine Learning is increasingly central in both industry and higher education, research on how to effectively teach it is still an emerging area. Shapiro and Fiebrink [26] highlight this gap and call for a dedicated research agenda for ML education, underlining the fact that ML education raises distinct research questions that go beyond traditional Computer Science (CS) pedagogy. They identify questions such as what types of mathematical, statistical, and computational prerequisites best prepare students for ML theory, and how learners conceptualize the structure and parameters of different ML algorithms.

Building on this need, several frameworks and practical strategies have been proposed. Allen et al. [1] introduced a framework for teaching Artificial Intelligence (AI) and Machine Learning in higher education, addressing the lack of systematic guidance on how these subjects should be taught effectively. Their work identifies key student learning barriers, particularly insufficient mathematical preparation and reduced confidence when dealing with abstract concepts. To address these challenges, they advocate for designing instruction around threshold concepts (topics in a curriculum where students typically get stuck) and for acknowledging diverse educational backgrounds. Similarly, Moosvi et al. [14] reviewed teaching practices from multiple experienced ML instructors across various institutions, in order to summarize them into common pedagogical strategies. Their work organizes effective teaching practices into thematic areas, such as preparing students with diverse mathematical and programming backgrounds, using real-world examples to motivate learning, and balancing intuition with technical content. Across these themes, they emphasize that successful ML education should not focus only on procedures, but also support conceptual understanding, critical thinking, and ethical awareness.

Complementing these pedagogical perspectives, Sahu et al. [24] investigated practical strategies for integrating ML concepts into existing undergraduate curricula, specifically within the field of Electrical Engineering. Their work integrated ML concepts directly into a "Signals and Systems" course, by relating them to core topics like orthogonality and dimensionality; for example, they introduced PCA in parallel with Fourier series, such that it was taught in context rather than as standalone material. The study showed mixed results: preliminary evidence suggested the integrated approach supports learning and is preferred by students, but numerical comparisons indicated that standalone workshop-based approaches used in prior work may lead to better performance.

2.2 The Mathematical Bottleneck

Many of the aforementioned works highlight the importance of the mathematical prerequisites necessary to understand most ML concepts. Tawfik et al. [32] investigated exactly which foundational skills actually predict student success in upper-level university ML courses by analyzing the grades, academic history, and prerequisite timing of students. They found that prerequisite grades do predict ML performance, but without being able to identify a "most important" prerequisite (e.g., Linear Algebra or Statistics). The work of Zhang and Allin [35] is closely related. The researchers introduced a "just-in-time" prerequisite review strategy, in which students were given targeted review questions and short instructional videos immediately before specific concepts were applied in a lecture. They found that this approach was well-received by students, helping to transfer prior math knowledge into the ML context and promoting self-regulated learning.

Sibia et al. [28] placed these findings in a generalized context by investigating the common challenges faced by learners taking introductory ML courses, and they identified barriers such as mathematical notation and vectorization, as well as challenges in conceptual understanding - most notably, students struggling with visualizing ideas and grasping the intuition behind them. The authors suggest that ML course design should include a combination of theory and practice, while also supporting students in understanding and using mathematical and ML notation.

2.3 Interactive Learning and Visualizations

Looking beyond the context of ML, the CS education community has also struggled with bridging the gap between abstract mathematical concepts and their algorithmic implementation. As a result, there has been an increasing shift toward learning environments that incorporate multiple representations and interactive elements. Shaffer et al. [25] argue for moving away from static reading and towards interactive textbooks in which students actively engage with the material through embedded simulations. Their engagement goals are grounded in the engagement taxonomy proposed by Naps et al. [15], who argue that increased engagement with learning materials leads to higher levels of learning within Bloom's taxonomy [3]. Rößling et al. [22] extend this idea by describing systems that integrate interactive visualizations directly into hypertextbooks, emphasizing that such visualizations should be embedded directly adjacent to their corresponding textual explanations. A well-known implementation of this approach is the OpenDSA system [5], which demonstrates both the technical feasibility and educational benefits of interactive e-textbooks in CS. Similarly, Moro [29] is a software tool specifically designed for AI and ML education that enables the integration of diverse types of materials - another step toward exploring multi-representational educational resources for ML instruction.

A particularly relevant foundation for this study is the recent work of Rențea et al. [20], who investigated the impact of static versus interactive visualizations in teaching ML concepts such as gradient descent and PCA to undergraduate students. Their interactive condition relied on Jupyter Notebooks containing 2D widgets that allowed students to manipulate specific parameters - such as adjusting the slope of a projection line to observe variance and reconstruction error, or manually scaling dataset dimensions to see the effect on principal components. In contrast, the interactive condition in the current study introduces a more comprehensive multimedia environment: linear algebra concepts are supported through narrated instructional videos with spatial visualizations, while a similar interactive widget allows students to dynamically observe variance and project data points onto the maximum variance line. Most significantly, the current study introduces a sequential 3D interactive widget for the PCA algorithm. Rather than interacting with isolated parameters, students can generate multiple random datasets and sequentially trigger each step of the PCA algorithm - mean centering, eigenvector calculation, finding principal components, and projecting onto lower dimensions - to observe the related geometric effects in real time (see Figure 1). Further explanations on instructional design considerations can be found in Section 3.2.

3 Methodology

The methodology followed during the study is described below.

3.1 Conceptual and Theoretical Background

Principal Component Analysis is a dimensionality reduction technique used to transform high-dimensional datasets into a lower-dimensional space while preserving as much variability in the data as possible. In practice, PCA is widely employed across different domains for feature extraction, noise reduction, and data visualization.

PCA begins by performing mean centering on the data; if features are measured on different scales, standardization is also needed to ensure compatibility. It then computes the covariance matrix to capture the linear relationships between variables. Through eigendecomposition, this matrix is factored to identify its eigenvectors and corresponding eigenvalues. The eigenvectors define a new set of orthogonal axes - the principal components - oriented in the directions of maximum data variance, while the eigenvalues quantify the amount of variance held along each axis. Finally, the original dataset is projected onto a selected subset of these top principal components, reducing its dimensionality while retaining its core informational structure.

3.2 Development of Instructional Materials

To isolate the effect of multiple representations on student comprehension, two distinct versions of a PCA learning module were developed using the Jupyter Book platform. [1]

The materials were structured into two experimental conditions:

- **Traditional Static Representations (Group A)**: Presented concepts using standard text combined with static images and formulas.
- **Multimedia-Enhanced Representations (Group B)**: Presented concepts using text and formulas, supplemented with narrated videos and interactive visualizations (e.g., an interactive 3D data graph).

A critical pedagogical design constraint was ensuring strict informational equivalence between the two versions; both books were designed to cover the same learning goals (see Table 1). The interactive version did not introduce supplementary concepts or additional knowledge. Instead, information was simply encoded in a different representation. The design of these materials was grounded in Cognitive Load Theory [31] and Mayer's Cognitive Theory of Multimedia Learning. [13]

Table 1: Learning Objectives for PCA Tutorial Books

Section	Learning Objective
Introduction to ML	Understand core machine learning concepts, including datasets, features, and feature vectors, and how real-world objects are represented numerically for computation.
Prerequisite Mathematics	Remember and interpret essential linear algebra and statistics concepts (vectors, matrices, variance, covariance) and relate them to their geometric interpretations in data space.
PCA Theory	Understand the full PCA pipeline, including mean-centering, covariance matrix construction, eigendecomposition, and projection onto principal components.
PCA Applied	Apply the PCA algorithm step-by-step to a complete numerical dataset to reinforce procedural understanding.

Specifically, the instructional design was grounded in Cognitive Load Theory by focusing on the primary goal of managing the intrinsic cognitive load associated with learning

[1] https://icvladareanu.github.io/v2-PCA-teaching-materials/

PCA. The content was structured according to a progressive scaffolding approach, in which students were gradually guided from foundational concepts to more complex ones. To further reduce extraneous cognitive load, explanatory text, formulas, and visualizations were spatially integrated in accordance with the spatial contiguity principle, with the goal of minimizing split-attention effects.

The PCA instructional materials were therefore organized into a sequence of five chapters. Chapter 1 introduced fundamental ML concepts, including features and feature vectors, datasets, and object representation. Chapter 2 provided a review of essential mathematical prerequisites, focusing on linear algebra and statistical concepts relevant to PCA. The motivation behind this chapter was to prepare students for transferring mathematical knowledge into the context of PCA, in line with Zhang and Allin's "just-in-time" prerequisite review strategy [35]. Chapter 3 introduced the PCA algorithm in a theoretical, step-by-step manner, while chapter 4 provided a fully worked-out numerical example in which each computational step was explicitly demonstrated to reinforce procedural understanding, consistent with the worked example principle. Chapter 5 concluded the tutorial with a short summary reinforcing the concepts taught.

In addition, the design of the Multimedia-Enhanced module (Group B) was informed by Mayer's Cognitive Theory of Multimedia Learning. In line with the dual-channel principle, narrated instructional videos were incorporated to distribute cognitive processing across auditory and visual channels. Furthermore, interactive 3D visualizations were included to support spatial reasoning in PCA by allowing students to directly manipulate data distributions in geometric space, as can be seen in Figure 1.

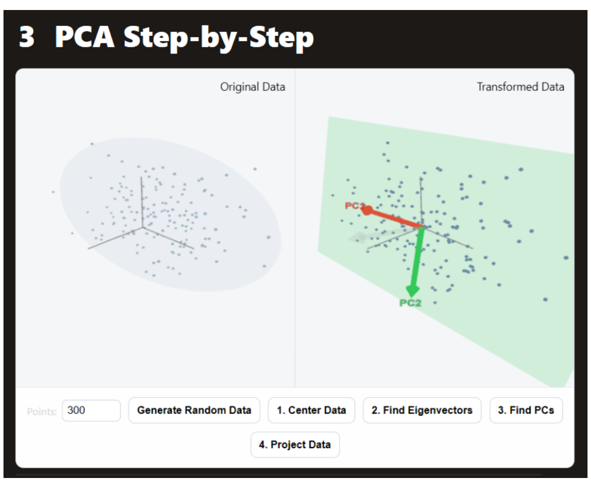

Figure 1: Interactive PCA widget showing an initial dataset on the left and the PCA-transformed dataset on the right

To ensure pedagogical and technical accuracy, the materials were reviewed by ML experts, including a teaching assistant, CS Master's and PhD students, and an assistant professor teaching the Machine Learning course at TU Delft. Due to project time constraints, the feedback was received and integrated into the learning materials only after the participant studies had already begun. Incorporating revisions at that stage would have introduced inconsistencies across experimental conditions, so only the original materials were used for the experimental study. A deeper overview of the specific issues identified by the experts can be found in Section 5.1.

3.3 Experimental Design

A between-subjects experimental design was used to compare the effects of different instructional representations. For an overview of the full methodology, see Figure 2.

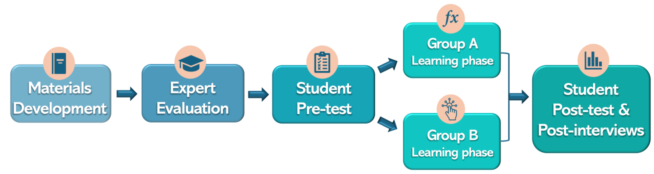

Figure 2: Methodology overview

Participants

Twenty-seven first-year Computer Science undergraduate students at TU Delft were recruited for the study. Participants were selected from this cohort because they had not yet taken any Machine Learning courses and were therefore not expected to have prior knowledge of PCA. However, as part of the CS curriculum, they were expected to have completed foundational courses in calculus, linear algebra, probability, and statistics.

Pre-test of Mathematical Knowledge

Prior to the learning phase, participants completed a baseline pre-test assessing their foundational mathematical knowledge in vectors, matrices, transformations, and basic statistics. Based on these pre-test scores, participants were assigned to either Group A (static) or Group B (interactive) using a matched-assignment strategy. This ensured that the average pre-test scores of the two groups were approximately equal, reflecting similar baseline knowledge of the prerequisite mathematics. The pre-test questions can be found in Appendix A.

Set-up and Environment

The experiment was conducted during on-campus sessions lasting 60 minutes in total. Participants were given 30 minutes for the learning phase, during which they independently read their assigned Jupyter Book tutorial on PCA. Immediately following this phase, they were given 20 minutes to complete a post-test. The last 10 minutes were reserved for completing a semi-structured post-interview.

Post-test of PCA Knowledge

The post-test measured PCA knowledge and consisted of multiple choice questions divided into three distinct sections, designed to evaluate varying depths of cognitive processing:

- **Conceptual Understanding**: Focused on evaluating the intuitive and logical comprehension of the PCA pipeline, ensuring students understood the "why" behind the steps rather than simply memorizing formulas.

- **Problem-Solving Performance**: Required students to apply small steps of the PCA algorithm to solve numerical mathematical exercises.
- **Knowledge Transfer**: Assessed the students' ability to apply PCA concepts to new, real-world contexts, such as image compression, medical research, and movie recommendation systems.

The post-test questions can be found in Appendix B.

Post-Interviews

Following the post-test, participants were asked to take part in a semi-structured interview to provide qualitative insights into their learning process. These reflection questions aimed to uncover their cognitive approaches to both the tutorial and the post-test. Questions were asked to pinpoint which specific representations triggered moments of understanding (when a concept "clicks"), and identify areas that caused extraneous cognitive load. Additionally, drawing inspiration from the Feynman technique, participants were asked to explain the concept of PCA in their own words, without using mathematical terminology - as if they were teaching it to a beginner. A full overview of the post-interview questions can be found in Appendix C.

Data Analysis

For the quantitative analysis, statistical tests were conducted on the pre-test and post-test results. The Shapiro–Wilk test was first used to assess the assumption of normality. When this assumption was violated, a non-parametric alternative - the Mann–Whitney U test - was applied. Separate Mann–Whitney U tests were conducted for each post-test category (Conceptual Understanding, Problem-Solving, and Knowledge Transfer). To account for multiple comparisons, a Bonferroni correction was applied.

The semi-structured interviews were analyzed using thematic analysis. A hybrid approach was used, as described by Fereday and Muir-Cochrane [4], combining deductive codes derived from theory with inductive codes emerging from the data. This approach is well-established in mixed-methods research [18] and has been previously applied in educational studies [34]. To enhance reliability and reduce individual bias, coding was performed by two independent coders, and inter-coder reliability (ICR) was calculated following the guidelines of Connor and Joffe [16]. After coding was completed, related codes were grouped into broader themes capturing recurring patterns across participants.

4 Results

The results are presented using both quantitative and qualitative analyses.

4.1 Quantitative Data

An overview of all statistical test results can be found in Table 2. A visual breakdown of each score per cognitive category can be seen in Figure 3. An overview of the pre-test scores (mathematical knowledge) and post-test scores (PCA knowledge) can be found in Figure 4. However, it should be noted that the tests evaluate different topics (mathematics versus ML knowledge) and should not be interpreted as a direct pre-post measure of learning gain.

Pre-test of Mathematical Knowledge

Normality of the pre-test scores was assessed using the Shapiro-Wilk test for both groups. The static group showed no violation of normality ($W = 0.8691$, $p = 0.0637$), while the interactive group showed a significant deviation from normality ($W = 0.8441$, $p = 0.0185$). Based on this violation, a non-parametric Mann-Whitney U test was used for group comparison.

The Mann-Whitney U test indicated no significant difference between the static ($M = 11.75$, $SD = 1.14$) and interactive group ($M = 11.71$, $SD = 1.20$) in pre-test scores ($U = 85$, $p = 0.9787$), as seen in the left-side graph in Figure 4.

Post-test of PCA Knowledge

Total scores. Normality testing showed that the interactive group violated the assumption of normality (static: $W = 0.7902$, $p = 0.0052$; interactive: $W = 0.8948$, $p = 0.095$). A Mann-Whitney U test was conducted and the results showed a statistically significant difference in post-test total scores between groups ($U = 141.5$, $p = 0.0111$), with the static group ($M = 8.31$, $SD = 0.75$) outperforming the interactive group ($M = 7.29$, $SD = 1.07$), as seen in the right-side graph in Figure 4.

Conceptual Understanding. Both groups showed non-normal distributions (static: $W = 0.5527$, $p < 0.001$; interactive: $W = 0.7502$, $p = 0.0013$). The Mann-Whitney U test with Bonferroni correction was applied, and showed no statistically significant difference between the static ($M = 2.62$, $SD = 0.77$) and interactive group ($M = 2.21$, $SD = 0.58$) in conceptual understanding ($U = 126.5$, $p = 0.0591$), as seen in the leftmost graph in Figure 3.

Problem Solving. Normality was violated in both groups (static: $W = 0.311$, $p < 0.001$; interactive: $W = 0.2968$, $p < 0.001$). A Mann-Whitney U test with Bonferroni correction was performed but no significant difference was found between the static ($M = 2.92$, $SD = 0.28$) and interactive group ($M = 2.93$, $SD = 0.27$) in problem solving ($U = 90.5$, $p = 1.000$), as seen in the middle graph in Figure 3.

Knowledge Transfer. Both groups were non-normally distributed (static: $W = 0.5332$, $p < 0.001$; interactive: $W = 0.7055$, $p < 0.001$), therefore a Mann-Whitney U test with Bonferroni correction was used. A statistically significant difference was found, with the static group ($M = 2.77$, $SD = 0.44$) scoring higher than the interactive group ($M = 2.14$, $SD = 0.77$) in knowledge transfer ($U = 136.5$, $p = 0.0131$), as seen in the rightmost graph in Figure 3.

4.2 Qualitative Data

Codebook Development. A comprehensive codebook consisting of 24 unique codes was developed to analyze the interview transcripts. The coding framework utilized a hybrid approach with both deductive and inductive codes. The deductive codes were theoretically driven, drawing primarily from Cognitive Load Theory [31] to identify instances of intrinsic, extraneous, and germane cognitive load, as well as from

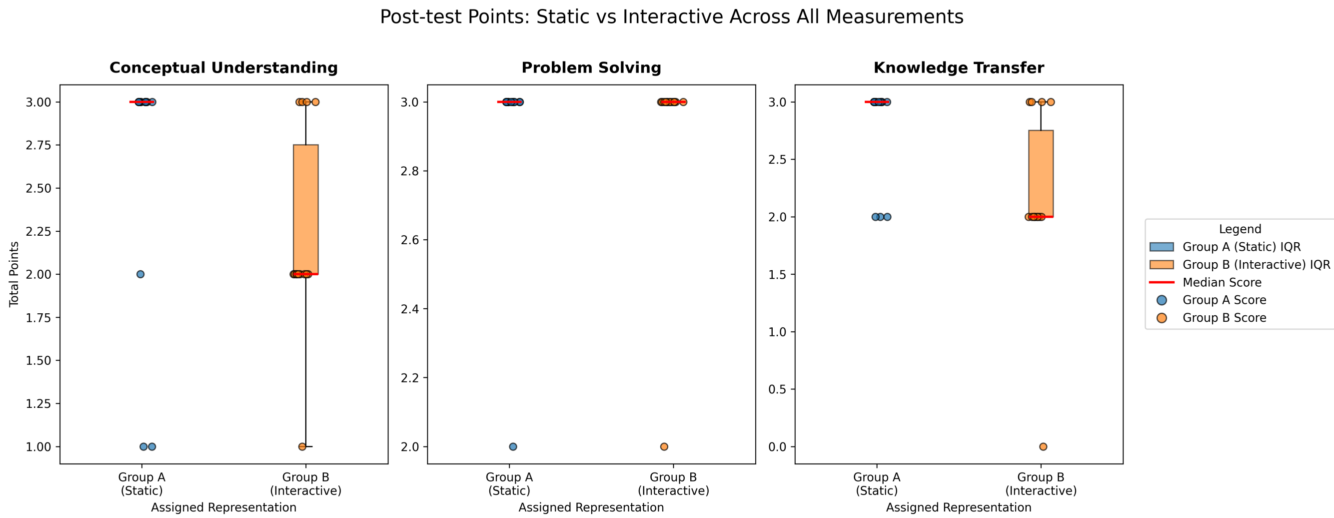

Figure 3: Post-test scores across the three distinct cognitive categories.

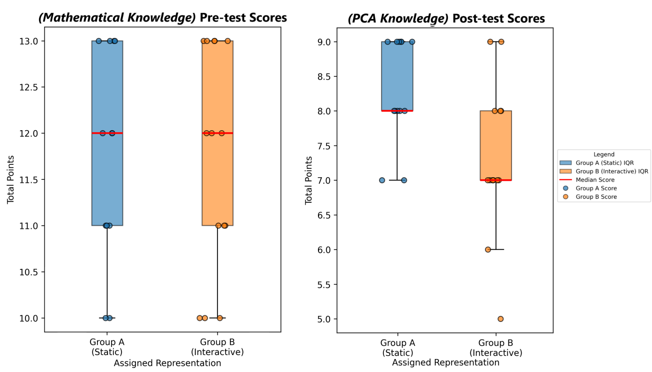

Figure 4: Pre-test and post-test total scores

the Cognitive Theory of Multimedia Learning [13] to capture phenomena such as contiguity, redundancy, and signaling. Inductive codes were generated directly from the participant interviews to capture specific recurring concepts, such as the students' reliance on geometric versus algebraic intuition, their familiarity with prior knowledge, and instances of misconception. A subset of four codes used in the analysis can be seen in Figure 5. The full codebook can be found in Appendix D.

Inter-Rater Reliability. Inter-rater reliability (IRR) was assessed to evaluate the consistency of coding between independent coders and ensure the reliability of the thematic analysis. A subset of interviews was chosen and coded independently by both coders; out of the 24 total codes in the codebook, 14 were actively used by both coders during this sample. The percentage agreement between coders was 82.5% ($p_o = 0.825$). To account for agreement occurring by chance, Cohen's Kappa (κ) was calculated, yielding a value of $\kappa = 0.7364$, indicating a substantial level of agreement [9].

Thematic Analysis. Following the coding process, a thematic analysis was conducted across all coded participant in-

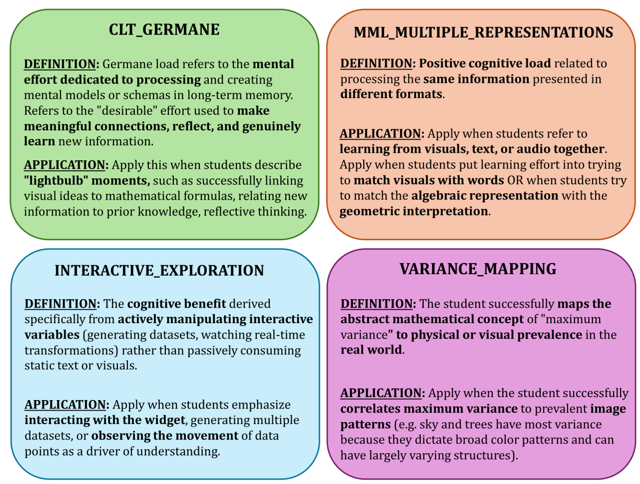

Figure 5: Subset of thematic analysis codes

terviews to identify broader patterns in the data. Nine distinct themes emerged; the full thematic observation overview can be found in Appendix E. A selection of relevant thematic areas will be discussed here.

First, most of the interactive group (11 out of 14) reported that the widgets consolidated their geometric intuition by allowing them to observe real-time transformations, as revealed by the theme "*Interactive Visualizations Facilitate Geometric Understanding of PCA*". Second, over half of all participants (14 of 27) felt the mathematical prerequisites were redundant because they had recently covered them in prior courses, as shown in the theme "*Prerequisite Materials are Perceived as Redundant If Recently Covered*". Third, geometric explanations generally supported deeper conceptual understanding, with 21 out of 27 participants noting that visual representations helped them decode abstract formulas - supported by the theme "*Geometric Representations Deepen Concep-*

Table 2: Overview of Statistical Test Results for Pre-test and Post-test Measures

	Static Group		Interactive Group		Mann–Whitney U	
Measure	M (SD)	SW W (p)	M (SD)	SW W (p)	U	p
Pre-test	11.75 (1.14)	0.8691 (0.0637)	11.71 (1.20)	0.8441 (0.0185)	85.0	0.9787
Post-test Total	8.31 (0.75)	0.7902 (0.0052)	7.29 (1.07)	0.8948 (0.0950)	141.5	**0.0111***
Conceptual Understanding	2.62 (0.77)	0.5527 (<0.001)	2.21 (0.58)	0.7502 (0.0013)	126.5	0.0591
Problem Solving	2.92 (0.28)	0.3110 (<0.001)	2.93 (0.27)	0.2968 (<0.001)	90.5	1.0000
Knowledge Transfer	2.77 (0.44)	0.5332 (<0.001)	2.14 (0.77)	0.7055 (<0.001)	136.5	**0.0131***

Note. SW = Shapiro–Wilk test. * Indicates statistical significance ($p < 0.05$).

p-values for Conceptual Understanding, Problem Solving, and Knowledge Transfer include a Bonferroni correction.

tual Understanding". However, observations of the theme "*Geometric Explanations and Interactive Features Were Not Equally Valuable for All Students*" also emerged: 3 participants in the interactive group reported the widgets as extraneous, while 2 in the static group preferred only algebraic text. Finally, although almost all participants could define PCA using standard terminology, only 9 out of 27 across both conditions successfully connected their explanations to deeper geometrical interpretations or real-world analogies, a disconnect reinforced by 11 participants (8 static, 3 interactive) admitting to guessing or making inaccurate assumptions during the post-test. This is supported by observations of the themes "*Students Struggled to Transfer PCA Concepts Beyond Memorized Definitions*" and "*Geometric and Multi-Representational Learning Supports Deeper Understanding of PCA*".

5 Discussion

This chapter interprets the results of the study and discusses their implications in light of several limitations.

5.1 Limitations

Before interpreting the results, several considerations must be acknowledged. First, the experimental study was conducted with a small sample size of 27 participants. It is possible that more subtle effects of interactive representations were present but went undetected due to a lack of statistical power. Second, the learning phase was strictly limited to 30 minutes. PCA is a dense algorithm, involving mean-centering, covariance matrices, eigendecomposition and dimensionality projection; as such, the 30 minute restriction introduced a significant potential confound - it is possible that the interactive group ran out of time. Because the 3D widgets were visually compelling, participants might have spent a disproportionate amount of time manipulating the simulations, leaving them with insufficient time to read or process the underlying text deeply, compared to the static group.

Furthermore, a limitation within the post-test design became apparent after analyzing the results of the Problem-Solving section. A clear ceiling effect occurred, where almost all participants scored full marks. This can be attributed to the post-test questions proving too easy, because they relied on procedural execution (plugging numbers into formulas) rather than more complex algorithm synthesis. Future studies should design more discriminating problem-solving questions, such as requiring students to perform the full PCA pipeline on 3D datasets, either in a partially unguided manner or with intermediate steps omitted.

Another limitation concerns the development of the instructional materials themselves. Although the materials were reviewed by ML experts, this feedback was received only after participant sessions had already begun; as such, incorporating revisions at that stage would have compromised the consistency of the experimental conditions. Consequently, the expert feedback was integrated a new version of the tutorial book, but it was not used in the study, making it possible that some of the identified issues in the instructional materials influenced the observed learning outcomes.

Reviewers identified that the original material lacked context, as it did not discuss limitations or real-life use cases of PCA - this may have limited participants' understanding of PCA relevance in a broader context. It was also noted that the connection between the statistical prerequisites and the PCA algorithm itself was left implicit, and that the static materials lacked intuitive explanations to situate PCA in a spatial or conceptual context. Finally, some figures intended to illustrate the intuition behind concepts were actually found to be confusing rather than helpful, because they were not mathematically accurate. To mitigate these issues, the improved materials introduced an additional chapter addressing both PCA limitations and practical applications. They also included intuitive examples and analogies for each step of the PCA algorithm, and replaced ambiguous figures with more mathematically accurate visualizations.

5.2 Overall Scores

The quantitative analysis revealed a statistically significant difference in post-test total scores, with the static representation group (Group A) outperforming the interactive group (Group B). However, a contradiction emerged when combining these statistics with the qualitative observations from the post-interviews.

During the interviews, most participants (across both conditions) verbally expressed a preference for the interactive and geometrical support, with considerably fewer students reporting these features as a source of extraneous cognitive load - as revealed by observations of the themes "*Geometric Representations Deepen Conceptual Understanding*" in combination with "*Geometric Explanations and Interactive Features Were Not Equally Valuable for All Students*". Yet, qualitative coding revealed that students in the static group were more likely to make inaccurate statements or admit to guessing on post-test answers. This suggests a possible disconnect: students' feeling of understanding and their preference for interactive media did not translate to objective performance gains. This could be attributed to the interactive design inadvertently introducing extraneous cognitive load, demanding too much working memory to operate the widgets while simultaneously processing the underlying mathematics. Alternatively, the widgets may have given the students a false sense of deep understanding, where the visual feedback made the PCA pipeline feel intuitive, but at the cost of reducing the germane cognitive effort students actually invested in internalizing the concepts.

5.3 Conceptual Understanding

When isolating conceptual understanding, the difference between the two groups was not statistically significant. Qualitative analysis supports this, as the theme "*Geometric and Multi-Representational Learning Supports Deeper Understanding*" was prevalent across both groups. This indicates that the core driver of conceptual understanding might not have been the interactivity itself, but rather the geometric narrative. The geometric explanations provided for both the mathematical prerequisites part, as well as for the PCA theory section, were well received by the majority of the students, with only a few individuals reporting them as redundant.

5.4 Problem-Solving Performance

No significant difference was observed between the groups in problem-solving performance. However, as noted in Section 5.1, a clear ceiling effect occurred across both conditions. Many participants reported prior familiarity with the mathematical concepts (as shown by observations of the theme "*Prerequisite Materials are Perceived as Redundant If Recently Covered*"), and because the post-test questions primarily required procedural execution (plugging numbers into formulas), this section may not have been sufficiently challenging to discriminate between the effects of the two instructional representations.

5.5 Knowledge Transfer

The knowledge transfer results were also a surprising finding. While post-interviews showed that students in both groups found it challenging to apply PCA concepts to unfamiliar real-world contexts, the static group performed significantly better than the interactive group.

One possible explanation is that the static materials required students to spend more time interpreting images or text and internalizing them. This may have encouraged a deeper engagement with the underlying principles of PCA rather than its visual representation alone; in the interactive group, the dynamic manipulation of variables provided immediate visual feedback, which may have come across only as transient information. While the interactive visualizations appeared to support immediate comprehension of geometric aspects of PCA, they did not necessarily guarantee that students connected what they observed to broader applications of the concept.

5.6 Implications for Educators

This study contributes several takeaways for ML educators and instructional materials designers. First, instructors should not assume that interactive materials will automatically improve learning; instead, careful attention is needed to ensure that interactive elements do not introduce extraneous cognitive load. Instructional design should further ensure that visualizations are paired with structured guidance, such as mandatory checkpoints, reflection questions, or guided tasks to ensure that students actively link visuals to the underlying algebraic formulations. Finally, if interactive elements are integrated, educators should proportionally adjust the allocated learning time. As observed in this study, under strict time constraints, interactive tools may become distracting, leading students to spend disproportionate time experimenting with the visualization rather than processing the deeper implications of core concepts.

6 Responsible Research

This chapter discusses the considerations made in order to assure that the research is conducted in a responsible, reproducible and ethical manner.

6.1 Ethical Considerations and Informed Consent

The experimental study was conducted in line with ethical research practices. Participation was fully voluntary, and the legal basis for processing participant data was the informed consent. All participants in the study were briefed on the study objectives, the data that would be collected and processed, and they were offered the right to withdraw from the experiment at any time, without consequences. The study did not pose any risk of mental fatigue to the participants, since only standard educational materials were used and evaluated.

6.2 Data Collection

The study collected administrative and research-specific data. Limited personally identifiable information (PII) - such as names and email addresses - was collected exclusively for administrative purposes, such as managing participation logistics and distributing the study details and instructional materials.

During the experiment, the study gathered minimal indirectly identifiable personal research data (PIIRD), such as prior educational background relevant to mathematics, as well as participant responses to the pre-test, post-test and post-interviews. Open reflection questions asked during the post-interview were used solely for qualitative reflection on the learning process; no sensitive personal data (health, gender, religion, etc.) was collected, and no audio, video or image recordings were taken during the sessions.

6.3 Data Processing and Anonymization

Data separation and anonymization were strictly enforced in order to protect participant privacy. The collected PII was not linked in any way to the research responses during data analysis. To prevent participants from being identified from their responses, all datasets were fully stripped of personal data and identifiable attributes, and qualitative and quantitative data analysis was conducted only on these fully anonymized datasets.

All collected data was stored securely, on university-approved storage systems with access restricted only to the research team. Only the primary researcher had access to the raw participant information containing personally identifiable data, which was stored and processed separately from the research-specific data.

6.4 Reproducibility

To ensure the reproducibility of the research and allow other research to validate the findings or reproduce the experiment, the final anonymized dataset will be archived. Personally identifiable information collected for administrative purposes was permanently deleted after the completion of the study. All pre-test, post-test and post-interview questions, and all versions of the instructional materials used in the study are documented and can be reused.

7 Conclusion

This study examined how different instructional representations affect students' learning of Principal Component Analysis. Although participants generally preferred the multimedia-enhanced materials and reported that the interactive visualizations helped them understand the geometric aspects of PCA, students who used the static materials achieved significantly higher overall post-test scores and performed better on knowledge transfer tasks. These findings suggest that interactive visualizations do not automatically improve learning outcomes. While they are highly engaging, they risk introducing extraneous cognitive load or creating an illusion of understanding if used passively. ML educators should carefully pair these visual tools with structured guidance to ensure students actively process the concepts, as well as making allowances for extended learning time.

Future Work. Several directions for future research emerge from this study. First, the experiment was limited to a short learning session; longer-term studies could provide a more realistic picture of how students learn complex ML concepts over time. Second, a larger and more diverse participant sample would improve statistical power and help identify if students actually benefit from interactive materials. Third, future iterations of the learning materials could build on the findings of this study to better balance interactivity and cognitive load. While participants found the interactive visualizations engaging and helpful for developing intuition, the results suggest that the way these tools were integrated may have reduced their effectiveness for deeper learning and knowledge transfer. Future designs could provide more structured guidance and explicitly connect visual interactions to the underlying mathematics.

A Pre-test Questions

Vectors and Geometry

1. Let $u_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $u_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Which coefficients produce $v = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$?
   - $c_1 = 1, c_2 = 2$
   - $c_1 = 2, c_2 = 1$
   - $c_1 = 3, c_2 = 1$
   - $c_1 = 4, c_2 = 0$
2. What is the norm of $v = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$?
   - 8
   - 5
   - 7
   - 25
3. What is the dot product of a and b? $a = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $b = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$
   - 1
   - 5
   - -1
   - -6
4. Are the following vectors orthogonal? $u = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $v = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$
   - Yes
   - No

Matrices and Transformations

5. Let $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$, $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Compute Av.
   - $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$
   - $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$
   - $\begin{pmatrix} 3 \\ 6 \end{pmatrix}$
   - $\begin{pmatrix} 1 \\ 6 \end{pmatrix}$
6. The matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ represents:
   - a 90 rotation
   - uniform scaling
   - a reflection across the x-axis
   - a reflection across the y-axis
7. Which transformation is not linear?
   - $T(x) = 3x$
   - $T(x) = Ax$
   - $T(x) = x + b$
   - $T(x) = 0$
8. Let $A = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$.

   a) Compute the characteristic polynomial $\det(A - \lambda I)$.
   - $\lambda^2 - 7\lambda + 14$
   - $\lambda^2 - 5\lambda - 10$
   - $\lambda^2 + 4\lambda + 7$
   - $\lambda^2 - 7\lambda + 10$

   b) Find the eigenvalues of A.
   - $\lambda = 7, 0$
   - $\lambda = 5, 2$
   - $\lambda = 4, 3$
   - $\lambda = 10, 3$

   c) Is the vector $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ an eigenvector of A?
   - Yes
   - No
9. Let A be a matrix and v be an eigenvector of A, such that $Av = \lambda v$. Which statement best describes what happens when A is applied to v?
   - The vector changes direction and length arbitrarily
   - The vector stretches or shrinks along its span
   - The vector becomes perpendicular to its original direction
   - The vector remains unchanged

Statistics

10. Given the data: $2, 4, 6, 8, 10$. What is the mean?
    - 5
    - 6
    - 8
    - 30
11. Given the data: $1, 3, 5$. What is the sample variance?
    - 2
    - 4
    - 8
    - 16

B Post-test Questions

Conceptual Understanding

1. A student is performing PCA on a dataset but forgets to perform the mean centering step. What is the most likely consequence when the algorithm calculates the first Principal Component (PC1)?
    - PC1 will correctly align with the direction of maximum variance, but all other principal components will be zero.
    - PC1 will point from the origin toward the center of the data cloud, capturing the dataset's physical location rather than its internal spread.
    - The principal components will remain completely unaffected because the covariance matrix automatically accounts for uncentered data.
    - The principal components will still point to the direction of the maximum variance, but their origin points will be shifted in space, starting from the center of the data cloud.
2. When PCA finds the first principal component (PC1) for a 2D dataset, it determines a line in the feature space. Which of the following properties uniquely defines this line?
    - It is the line that minimizes the sum of the absolute horizontal and vertical distances from the data points to the line.
    - It maximizes the variance of the datapoints when they are projected onto the line, which is the same as minimizing the perpendicular distances from the datapoints to the line.
    - It connects the data point with the lowest values directly to the data point with the highest values.
    - It is the line that passes through the origin and ensures that the total distance between all pairs of projected points is minimized.
3. The last step of the PCA algorithm - Projecting the Data - involves multiplying the centered dataset by the matrix of eigenvectors W. Conceptually, what does this matrix multiplication do to the data points?
    - It rotates the dataset so that the original feature axes align with the directions of greatest variance.
    - It rescales each feature so that features with larger variance contribute more strongly to the final representation.
    - It reconstructs the dataset using only the eigenvectors, replacing the original feature values with computed eigenvalues.
    - It projects the original data points onto the newly established orthogonal axes, effectively reducing the number of dimensions.

Problem Solving

4. **Problem Statement:** You have calculated the covariance matrix for a dataset to be:

    $$C = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

    Set up the characteristic equation $\det(C - \lambda I) = 0$ and solve for the eigenvalues λ_1 and λ_2.
    - $\lambda_1 = 3, \lambda_2 = 1$
    - $\lambda_1 = 1, \lambda_2 = 1$
    - $\lambda_1 = 2, \lambda_2 = 3$
    - $\lambda_1 = 3, \lambda_2 = 2$
5. **Problem Statement:** Consider the covariance matrix:

    $$C = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

    The eigenvalues of this matrix are given as: $\lambda_1 = 4$, $\lambda_2 = 2$. Find the corresponding eigenvectors and normalize them.

    Hint 1: An eigenvector v satisfies: $(C - \lambda I)v = 0$

    Hint 2: A normalized vector has length $= 1$. To normalize a vector v, you can divide it by its norm, where the norm is: $||v|| = \sqrt{v_1^2 + \cdots + v_n^2}$
    - A) $v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$
    - B) $v_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, v_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}$
    - C) $v_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, v_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}$
    - D) $v_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, v_2 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$
6. **Problem Statement:** Consider the following perfectly centered data matrix X, where the rows represent 3 distinct samples and the columns represent 2 features:

    $$X = \begin{pmatrix} 1 & 7 \\ -7 & 1 \\ 6 & -8 \end{pmatrix}$$

    We have found the following eigenvalues and their corresponding normalized eigenvectors:
    - $\lambda_1 = 150$, $\lambda_2 = 50$
    - $w_1 = \begin{pmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{pmatrix}, w_2 = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix}$

    Your Task: Apply Step 4 and Step 5 of the Principal Component Analysis pipeline to reduce the dimensionality of this dataset from 2D to 1D. Find the final projected data matrix Y.

    Reminder: Step 4 (Finding the PC): Select the Principal Component corresponding to the correct eigenvalue. Step 5 (Projecting the Data): Map the 2D data into the 1D space using the projection formula $Y = X_{\text{centered}}W$
    - A) $Y = \begin{pmatrix} 5 \\ 5 \\ -10 \end{pmatrix}$
    - B) $Y = \begin{pmatrix} 10 \\ 0 \\ -5 \end{pmatrix}$

- C) $Y = \begin{pmatrix} -5 \\ -5 \\ 10 \end{pmatrix}$
- D) $Y = \begin{pmatrix} 5 \\ -5 \\ -10 \end{pmatrix}$

Knowledge Transfer

7. If you use PCA to compress a large, highly detailed photograph of a forest into a smaller file size, what part of the image does the first principal component capture?
   - The random static or digital noise in the background.
   - The overall ”big picture” structures, like the broad colors of the sky, ground, and tree trunks.
   - The tiny details, like the individual leaves or flowers on the ground.
   - The metadata of the picture, like the file name and timestamp of when the photo was taken.
8. A streaming service tracks 100 different habits about how you watch movies (time of day, genres you like, etc.), representing each user as a data point in a 100-dimensional feature space. To optimize their recommendation engine, engineers use PCA to project this dataset down into a 10-dimensional space. What is the main trade-off they are making?
   - The system will run faster, but it might miss subtle, specific quirks about your personal taste.
   - They save processing power by identifying and deleting the 90 least important original tracking metrics, keeping only the 10 most predictive habits.
   - The app will now only be able to recommend movies from exactly 10 specific genres.
   - The system will run faster, but it will recommend less movies than before.
9. Medical researchers have a dataset of 500 patients. For each patient, they measure 10,000 different genetic markers. They want to know if the patients are ”Healthy” or ”At-Risk”. Why would the researchers use PCA in this scenario?
   - PCA will automatically search through the 10,000 markers, identify the single gene causing the disease, and delete the rest.
   - PCA increases the variance of the healthy patients so they stand out more clearly from the at-risk patients.
   - PCA will naturally group the patients into exactly two groups (”Healthy” or ”At-Risk”).
   - Humans cannot visualize a 10,000-dimensional graph, but PCA can condense the 10,000 variables into 2 or 3 dimensions, making it possible to plot the data on a standard scatterplot.

C Post-interview Questions

0. What group were you assigned to at the beginning of the session?
1. Was there a specific part of the tutorial – whether it was a block of text, an explanation, a visual, or a widget – that made the concept ‘click’ for you?
2. Was there an explanation or a widget or video that you felt you could have skipped entirely without losing any understanding? Did it add more cognitive load with no gain?
3. Did you use the algebraic text/formula to understand the geometric explanation/visualization, or did the geometric explanation/visualization help you understand the algebraic text/formula better?
4. If you had to explain the mechanism/goal of PCA to another first-year student who hasn’t taken this tutorial, how would you describe it without using any math terms?
5. Can you think of a scenario – outside of the examples provided in the tutorial/post-test – where PCA would be useful?
6. Let’s look back at one of the applied scenario questions you answered in the post-test. Can you walk me through your thought process when you selected your answer? Can you try to explain the process through which you arrived at this answer, or discarded the other ones?

   [If you use PCA to compress a large, highly detailed photograph of a forest into a smaller file size, what part of the image does the first principal component capture?]
   - a) The overall ”big picture” structures, like the broad colors of the sky, ground, and tree trunks.
   - b) The random static or digital noise in the background.
   - c) The tiny details, like the individual leaves or flowers on the ground.
   - d) The metadata of the picture, like the file name and timestamp of when the photo was taken.
7. Think back to the 5-step process of PCA. Could you walk me through your conceptual understanding of this process? For each step, what is the first thing that comes to mind? Do you immediately picture a mathematical formula, or do you picture a visual dataset, e.g., eigenvectors represented in vector space? Can you picture what the data would look like (e.g., point clouds in a graph) and what happens to it after each step?
   - Step 1: Centering the Data
   - Step 2: Calculating the Covariance Matrix
   - Step 3: Eigenvalue Decomposition
   - Step 4: Finding Principal Components
   - Step 5: Project the Data

D Codebook

Table 3: Complete Interview Codebook (Part 1: General Codes)

Code Name	Type	Definition	Application in Data	Source
CLT_Intrinsic	Deductive	The intrinsic difficulty or complexity of new information to be processed – i.e. the level of difficulty of a new task or topic.	Apply when students discuss the unavoidable complexity of topics such as dimensionality reduction, eigenvectors, or covariance structures.	Cognitive Load Classification by Sweller, van Merriënboer, and Paas (1998)
CLT_Extraneous	Deductive	Extraneous load results from the way information is presented to students and is not directly related to the concept itself. It is generated by poor instructional design, unnecessary complexity, or irrelevant information that interferes with the learning process.	Apply this when students mention confusing UI or navigation, dense text, irrelevant information or images.	Cognitive Load Classification by Sweller, van Merriënboer, and Paas (1998)
CLT_Germane	Deductive	Germane load refers to the mental effort dedicated to processing and creating mental models or schemas in long-term memory. Refers to the "desirable" effort used to make meaningful connections, reflect, and genuinely learn new information.	Apply this when students describe "lightbulb" moments, such as: successfully mentally linking visual ideas to mathematical formulas, relating new information to prior knowledge, reflective thinking or self-explanation of concepts.	Cognitive Load Classification by Sweller, van Merriënboer, and Paas (1998)
MML_Contiguity	Deductive	The cognitive benefit of keeping related text, math, and visuals physically close together or presenting them at the same time.	Apply this when students discuss the layout of the tutorial: text placed next to the corresponding interactive widget, saving them from having to scroll back and forth, or seeing theoretical text with calculations or examples presented in the same space.	Cognitive Theory of Multimedia Learning, by Mayer (2001)
MML_Redundancy	Deductive	Cognitive effort caused by unnecessary duplication of information in multimedia materials. Students experience overload when the same content is presented in multiple forms (e.g., spoken and written text simultaneously) without adding value.	Apply this when students mention feeling overwhelmed, distracted or confused by repeated information, such as "too much text," "same thing in video and text," or "explanations repeated unnecessarily."	Cognitive Theory of Multimedia Learning, by Mayer (2001)
MML_Multiple_Representations	Deductive	Positive cognitive load related to processing the same information presented in different formats.	Apply when students refer to learning from visuals, text, or audio together. Apply when students put learning effort into trying to match visuals with words OR when students try to match the algebraic representation with the geometric interpretation.	Cognitive Theory of Multimedia Learning, by Mayer (2001)
MML_Signaling	Deductive	Signaling refers to the use of visual cues to highlight essential information and guide the student's attention to the most critical information.	Apply when students refer to highlights, bold text, or dropdowns that guided their focus.	Cognitive Theory of Multimedia Learning, by Mayer (2001)
MML_Worked_Example	Deductive	Refers to when people learn better when they receive worked examples in initial skill learning: using step-by-step solved examples to support learning by demonstrating problem-solving strategies.	Apply when students mention understanding a task because they see a worked example.	Cognitive Theory of Multimedia Learning, by Mayer (2001)
ML_Geometric_Intuition	Inductive	Refers to understanding the physical/geometric perspective of a mathematical concept.	Apply when students describe the geometric explanations helping them visualize mathematical concepts such as centering, vector projections, data spread or dimensionality reduction.	Participant Interviews
ML_Algebraic_Intuition	Inductive	Refers to understanding a mathematical concept purely from the formula or theoretical explanation.	Apply when students mention viewing the concept purely through mathematical formulas, or when students focus on the algebraic steps.	Participant Interviews
Prior_Knowledge_Familiarity	Inductive	Refers to when students skip over introductory/prerequisite materials (like Linear Algebra or Statistics) because they have recently covered these topics.	Apply when students mention skipping text, videos, or formulas because the concepts were fresh from a recent course.	Participant Interviews
Interactive_Exploration	Inductive	The cognitive benefit derived specifically from actively manipulating interactive variables (generating datasets, watching real-time transformations) rather than passively consuming static text or visuals.	Apply when students emphasize interacting with the widget, generating multiple datasets, or observing the movement of data points as a driver of understanding.	Participant Interviews
Misconception	Inductive	The explanation of the student contains a fundamental flaw in how the algorithm works.	Apply this when the student gives an inaccurate, wrong or nonsensical explanation.	Participant Interviews

Table 4: Complete Interview Codebook (Part 2: Question-Specific Codes)

Code Name	Type	Definition	Application in Data	Source
Question 4: PCA Explain using Feynmann's Technique				
Feyn_Dimensionality	Inductive	The student explicitly mentions the reduction of space, size, or number of variables.	Apply when the student explicitly mentions reducing the data size, the number of features, etc. Also applicable when students mention having complex or highly dimensional datasets.	Participant Interviews
Feyn_Variance	Inductive	The student correctly identifies that the algorithm actively searches for and retains the "most important" or most varied information or mentions reducing correlated or redundant features.	Apply this when the student explicitly mentions that PCA captures the features with the most variance and retains them.	Participant Interviews
Feyn_Geometric	Inductive	The student uses clear, spatial, or structural language to explain the transformation without relying on explicit math formulas. They describe the geometrical effect of PCA accurately.	Apply this when the student explicitly describes geometric effects of PCA, such as projecting data or finding new axes.	Participant Interviews
Feyn_Analogy	Inductive	The student relies on informal, non-geometric or non-mathematical metaphors or storytelling to explain the concept.	Apply this when the student explains PCA using analogies for the dataset or PCA steps.	Participant Interviews
Feyn_Misconception	Inductive	The explanation of the student contains a fundamental flaw in how the algorithm works.	Apply this when the student gives an inaccurate, wrong or nonsensical explanation.	Participant Interviews
Question 5: PCA Application Scenarios				
High_Dimensions	Inductive	The student identifies a scenario explicitly driven by a massive number of features.	Apply when the student demonstrates that they understand *when* PCA is necessary, in the case of high dimensions.	Participant Interviews
Correlation	Inductive	The student explicitly mentions finding patterns in the data or reducing redundant features within their scenario.	Apply when the student demonstrates that they understand *when* PCA is necessary, in the case of redundant features.	Participant Interviews
Superficial	Inductive	The student names a generic "big dataset" without articulating why PCA is the right tool.	Apply when the student shows only a superficial understanding of PCA.	Participant Interviews
Question 6: Cognitive Process				
Variance_Mapping	Inductive	The student successfully maps the abstract mathematical concept of "maximum variance" to physical or visual prevalence in the real world.	Apply when the student successfully correlates maximum variance to prevalent image patterns (e.g. sky and trees have most variance because they dictate broad color patterns).	Participant Interviews
Goal_PCA	Inductive	The student argues backward from the goal of PCA.	Apply when the student motivates their answer by quoting the goal of PCA (e.g. "PCA keeps the most important things, and the big picture is the most important").	Participant Interviews
Elimination_Guess	Inductive	The student arrived at the answer purely by discarding other options or explicitly admitted to guessing the answer.	Apply this when the student admits to guessing the answer or doing purely elimination strategies.	Participant Interviews

E Identified Themes

Table 5: Summary of Identified Themes and Corresponding Codes

Theme Title	Observations per Group	Codes	Relevant Quotes
Integrated Instructional Design Supports Conceptual Understanding	Observed equally across both groups	MML_Signaling, MML_Contiguity, MML_Worked_Example, CLT_Germane	• "...the keywords and highlighted text for the most important concepts; those were useful to have. The numbering of steps gave a clear and logical overview..." • "...the way the text was arranged - especially for some concepts, there was a dropdown with an extra explanation or a block of highlighted text..."
Interactive Visualizations Facilitate Geometric Understanding of PCA	Observed across 11 out of 14 participants from the interactive group	ML_Geometric_Intuition, Interactive_Exploration, CLT_Germane, MML_Multiple_Representations	• "I think if I only had the graph and the simulations and the videos, I could have skipped all the theory." • "I really liked that for all of them, I could generate the dataset, and I could see how the transformations occurred..."
Prerequisite Materials are Perceived as Redundant If Recently Covered	Observed across 14 out of 27 participants from both groups	CLT_Extraneous, Prior_Knowledge_Familiarity, MML_Redundancy, MML_Multiple_Representations	• "The first two videos that explained linear algebra felt useless because the concepts were already covered in a recent linear algebra course that we took." • "The mathematical prerequisites were already covered, so they felt redundant."
Geometric Representations Deepen Conceptual Understanding	Observed across 21 out of 27 participants from both groups	ML_Geometric_Intuition, MML_Multiple_Representations, CLT_Germane	• "I understood the formulas, but the geometric visualizations really helped me see what those formulas were actually doing." • "I think visual explanations, so geometric ones, help a lot - especially for me, I am a visual learner..."
Worked Examples Bridge Theory and Application	Observed across 10 out of 27 participants from both groups	MML_Worked_Example, CLT_Germane	• "For the PCA part, the example helped me in understanding how to do the actual PCA..." • "The theory of the 5 PCA steps was not as useful as seeing the applied PCA example. It was much easier to learn from the practical one"
Geometric Explanations and Interactive Features Were Not Equally Valuable for All Students	Three (3) participants from the interactive group reported the interactive widget as extraneous cognitive load. Two (2) participants from the static group reported the geometric explanations as redundant.	CLT_Extraneous, ML_Geometric_Intuition, Interactive_Exploration	• "...the interactive widget helped me see what was happening, but I could have skipped it and I still would have understood PCA." • "I think the geometric explanation was just additional – for me the words communicate more than visuals..."
Students Struggled to Transfer PCA Concepts Beyond Memorized Definitions	Observed 8 static group participants making inaccurate statements or guessing for question 6. Observed 3 of the same in participants from the interactive group.	Misconception, ML_Algebraic_Intuition, Superficial, Goal_PCA, Elimination_Guess	• "I don't remember what answer I gave – I was guessing for this exercise. I couldn't picture how PCA would be applied to an image." • "...the biggest variance would be in the smallest details, which separate one forest from another."
Using Multiple Representations Helps Connect Variance to Real-World Data Interpretation	Observed across 8 participants from both groups.	MML_Multiple_Representations, Variance_Mapping	• "I just thought that since the biggest things in the picture like the sky or the trees are the most important features, we want to keep them." • "I remembered from the tutorial that the first principal component remembers the most variance..."
Geometric and Multi-Representational Learning Supports Deeper Understanding of PCA	Almost all participants could explain PCA using terminology related to dimensionality or variance. However, only 9 out of 27 connected their explanation to deeper geometrical interpretations or other analogies.	ML_Geometric_Intuition, MML_Multiple_Representations, Feyn_Dimensionality, Feyn_Variance, Feyn_Geometric, Feyn_Analogy	• "Imagine we are trying to map out something in the world, like dinosaurs... so we want to reduce the amount of these different variables to a workable number." • "Doing PCA is like taking a dataset with a lot of information, then picking out the directions in which that information is most important."

References

[1] Becky Allen, Andrew Stephen McGough, and Marie Devlin. Toward a framework for teaching artificial intelligence to a higher education audience. *ACM Transactions on Computing Education*, 22(2):1–29, 2021.

[2] Robert K. Atkinson. Multimedia Learning of Mathematics. In Richard Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, pages 393–408. Cambridge University Press, Cambridge, 2005.

[3] B. S. Bloom, M. B. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl. *Taxonomy of educational objectives. The classification of educational goals. Handbook 1: Cognitive domain*. Longmans Green, New York, 1956.

[4] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1):80–92, 2006.

[5] Eric Fouh, Ville Karavirta, Daniel A. Breakiron, Sally Hamouda, Simin Hall, Thomas L. Naps, and Clifford A. Shaffer. Design and architecture of an interactive etextbook – the opendsa system. *Science of Computer Programming*, 88:22–40, 2014.

[6] Michael Greenacre, Patrick Groenen, Trevor Hastie, Alfonso Iodice D'Enza, Angelos Markos, and Elena Tuzhilina. Principal component analysis. *Nature Reviews Methods Primers*, 2:100, 2022.

[7] Adedamola A. Kareem. The use of multimedia in teaching biology and its impact on students' learning outcomes. *The Eurasia Proceedings of Educational and Social Sciences*, 9:157–165, 2018.

[8] R. Kozma and J. Russell. Multimedia learning of chemistry. In R. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, pages 409–428. Cambridge University Press, Cambridge, 2005.

[9] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[10] Nathan P. Lawrence, Seshu Kumar Damarla, Jong Woo Kim, Aditya Tulsyan, Faraz Amjad, Kai Wang, Benoit Chachuat, Jong Min Lee, Biao Huang, and R. Bhushan Gopaluni. Machine learning for industrial sensing and control: A survey and practical perspective. *Control Engineering Practice*, 145:105841, 2024.

[11] W. Li and T. Li. Comparison of deep learning models for predictive maintenance in industrial manufacturing systems using sensor data. *Scientific Reports*, 15:23545, 2025.

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

[13] Richard E. Mayer. Multimedia learning. *Psychology of Learning and Motivation*, 41:85–139, 2002.

[14] Firas Moosvi, Fraida Fund, Varada Kolhatkar, Meiying Qin, Thomas Price, and Lisa Zhang. Effective strategies for teaching machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 40661–40669, 2026.

[15] Thomas L. Naps, Guido Rößling, Vicki Almstrum, Wanda Dann, Rudolf Fleischer, Chris Hundhausen, Ari Korhonen, Lauri Malmi, Myles McNally, Susan Rodger, and J. Ángel Velázquez-Iturbide. Exploring the role of visualization and engagement in computer science education. *SIGCSE Bulletin*, 35(2):131–152, 2002.

[16] Cliodhna O'Connor and Helene Joffe. Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19:1609406919899220, 2020.

[17] J. L. Plass and L. C. Jones. Multimedia learning in second language acquisition. In R. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, pages 467–488. Cambridge University Press, Cambridge, 2005.

[18] Kevin Proudfoot. Inductive/deductive hybrid thematic analysis in mixed methods research. *Journal of Mixed Methods Research*, 17(3):308–326, 2023.

[19] Viktorija Pruckovskaja, Axel Weissenfeld, Clemens Heistracher, Anita Graser, Julia Kafka, Peter Leputsch, Daniel Schall, and Jana Kemnitz. Federated learning for predictive maintenance and quality inspection in industrial applications, 2023.

[20] Ilinca Renţea, Gosia Migut, and Jesse Krijthe. Are interactive visualizations in machine learning education helping students? In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, ITiCSE 2025, pages 2–8, New York, NY, USA, 2025. Association for Computing Machinery.

[21] J. F. Rodrigues, L. Florea, M. C. F. de Oliveira, et al. Big data and machine learning for materials science. *Discover Materials*, 1:12, 2021.

[22] Guido Rößling, Thomas Naps, Mark S. Hall, Ville Karavirta, Andreas Kerren, Charles Leska, Andrés Moreno, Rainer Oechsle, Susan H. Rodger, Jaime Urquiza-Fuentes, and J. Ángel Velázquez-Iturbide. Merging interactive visualizations with hypertextbooks and course management. *SIGCSE Bulletin*, 38(4):166–181, 2006.

[23] Edoardo Saccenti. A gentle introduction to principal component analysis using tea-pots, dinosaurs, and pizza. *Teaching Statistics*, 46:38–52, 2024.

[24] Chinmay Sahu, Blaine Ayotte, and Mahesh K. Banavar. Integrating machine learning concepts into undergraduate classes. In *2021 IEEE Frontiers in Education Conference (FIE)*, pages 1–5. IEEE, 2021.

[25] Clifford A Shaffer, Thomas L Naps, and Eric Fouh. Truly interactive textbooks for computer science educa-

tion. In *Proceedings of the Sixth Program Visualization Workshop*, pages 97–103, 2011.

[26] R. Benjamin Shapiro and Rebecca Fiebrink. Introduction to the special section: Launching an agenda for research on learning machine learning. *ACM Transactions on Computing Education*, 19(4):1–6, 2019.

[27] Jonathon Shlens. A Tutorial on Principal Component Analysis. *arXiv*, 2014.

[28] Naaz Sibia, Amber Richardson, Alice Gao, Andrew Petersen, and Lisa Zhang. Student perspectives on the challenges in machine learning. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, pages 9–15, New York, NY, USA, 2025. Association for Computing Machinery.

[29] Sameer Singh and Sebastian Riedel. Creating interactive and visual educational resources for ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[30] M. S. Smith and D. Lu. Machine learning opportunities for nucleosynthesis studies. *Frontiers in Astronomy and Space Sciences*, 11:1494439, 2024.

[31] John Sweller, Jeroen J. G. van Merrienboer, and Fred G. W. C. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296, 1998.

[32] Marina Tawfik, Andrew Petersen, Leo Porter, and Lisa Zhang. Prerequisites and performance in a machine learning course: A quantitative analysis. In *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, New York, NY, USA, 2025. Association for Computing Machinery.

[33] J. Vamathevan, D. Clark, P. Czodrowski, et al. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18:463–477, 2019.

[34] Wen Xu and Katina Zammit. Applying thematic analysis to education: A hybrid approach to interpreting data in practitioner research. *International Journal of Qualitative Methods*, 19:1609406920918810, 2020.

[35] Lisa Zhang and Sonya Allin. Just-in-time prerequisite review for a machine learning course. In *Proceedings of the 25th Western Canadian Conference on Computing Education*, New York, NY, USA, 2023. Association for Computing Machinery.