Teaching Principal Component Analysis Through Multiple Representations

Impacts on Conceptual Understanding, Problem Solving and Knowledge Transfer

Author: Iulia-Cristiana Vlădăreanu (I.Vladareanu@student.tudelft.nl) | Responsible Professor: Gosia Migut | Supervisor: Ilinca Rențea

1 Introduction

The rapid integration of **Machine Learning (ML)** into industrial applications and academic research has led to a substantial increase in ML courses in undergraduate and graduate STEM curricula. As ML becomes a foundational competence for engineers and scientists, the quality of ML education is increasingly crucial.

ML relies on **abstract mathematical foundations**. Core topics require knowledge from multivariable calculus, linear algebra and high-dimensional geometry. Students must coordinate algebraic manipulation, geometric intuition, and algorithmic thinking simultaneously. This creates a major learning challenge for students with limited mathematical intuition.

Background

Principal Component Analysis (PCA) is a widely used **dimensionality reduction** technique. It reduces data complexity while preserving features in a dataset.

Literature shows how PCA can be taught using **different perspectives:**

- **Linear algebra** and **Statistics** → Covariance matrices, eigenvalue decomposition, variance decomposition, standardization [1] [2]
- **Analogies** → Concrete examples and intuitive analogies [3]
- **Visualizations** → Interactive visualization tools [4] [5] [6]

Motivation

Research in **multimedia learning** shows that understanding improves when concepts are taught through multiple representations. [7] [8] However, ML education lacks empirical evidence on how **multi-representational instruction** affects the learning process of concepts such as PCA.

2 Research Question

How does the combination of multiple instructional representations across the full learning pipeline - from prerequisite mathematical concepts to PCA instruction - affect student learning?

- **RQ1:** How does it affect conceptual understanding?
- **RQ2:** How does it affect problem-solving performance?
- **RQ3:** How does it affect knowledge transfer ability?

3 Methodology

Materials Development → **Expert Evaluation** → **Student Pre-test** → **Group A** (Learning phase: static condition) / **Group B** (Learning phase: interactive condition) → **Student Post-test & Post-interviews**

Materials Development:
- A: Traditional Static Representations - text, static images
- B: Multimedia-Enhanced Representations - text, videos, interactive widgets

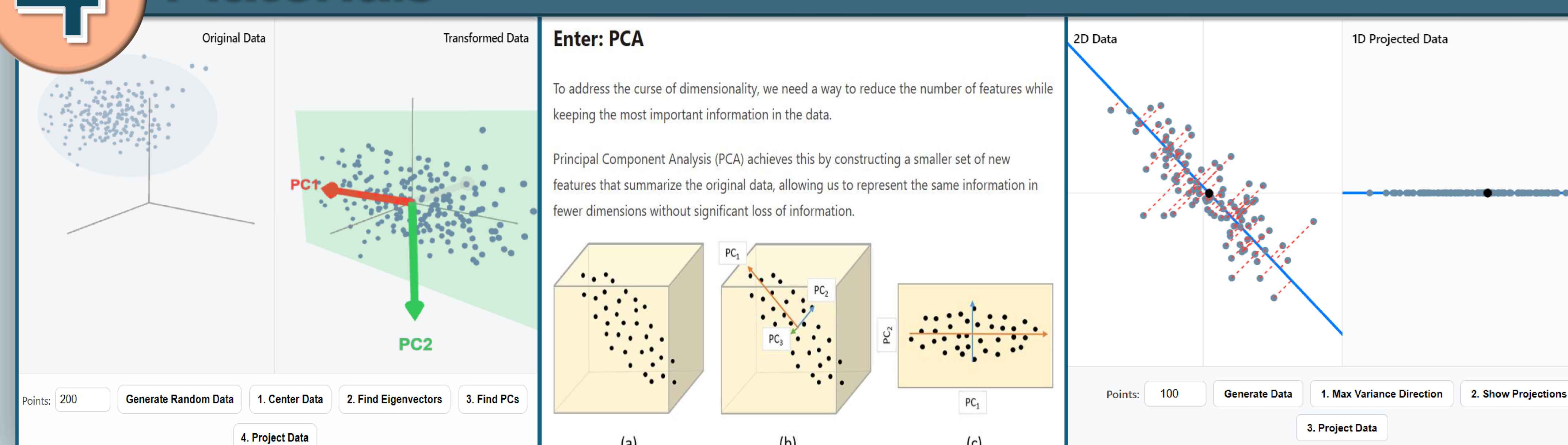
Expert Evaluation:
- **ML experts**: TA, CS Masters student, CS PhD student, ML assistant professor
- Collect **feedback** and refine materials

Student Pre-test:
- Participant group: n = 27
- Tests prior **mathematical knowledge**
- Split participants into 2 balanced groups with **equal average scores**

Group A / Group B:
- 30-min. in-person **learning phase**
- Participants complete learning phase using **assigned materials**

Student Post-test & Post-interviews:
- **Quantitative data**:
  - Understanding
  - Problem-solving
  - Knowledge transfer
- **Qualitative data**:
  - Thought process
  - Learning insights

4 Materials



5 Results

CLT_GERMANE
Mental effort "desirable" for making **connections, reflecting, learning** new information.

CLT_EXTRANEOUS
Difficulty or complexity of new information to be processed.

MML_MULTIPLE_REPRESENTATIONS
Positive cognitive load - processing the **same information** presented in **different formats:** visuals, text, or audio together.

MML_WORKED_EXAMPLE
When **people learn better** when they receive step-by-step worked examples.

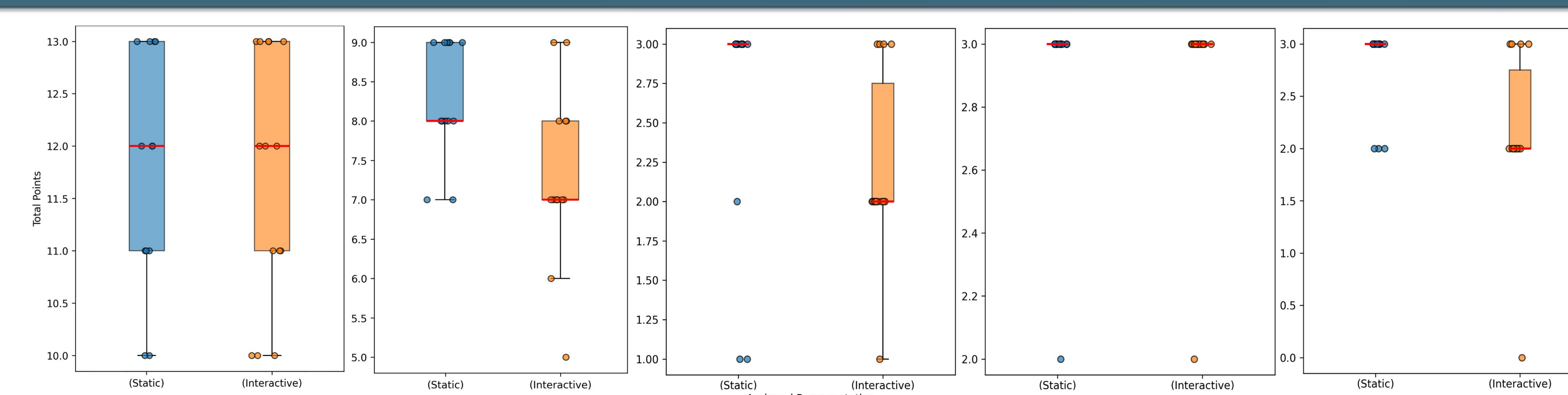
INTERACTIVE_EXPLORATION
Cognitive benefit from **actively manipulating interactive variables.**

PRIOR_KNOWLEDGE_FAMILIARITY
When students **skip over prerequisite materials** because they have recently covered the topics.

VARIANCE_MAPPING
Student successfully **maps the abstract mathematical concept** of "maximum variance**" to visual prevalence** in the **real world**.

FEYN_GEOMETRIC
Student uses clear, **spatial, or structural language to explain the transformation** without relying on explicit math formulas.

Codebook (subset) developed for Hybrid Thematic Analysis



Pre-test | Post-test: total scores | Post-test: understanding | Post-test: problem-solving | Post-test: transfer

6 Discussion

Static group **outperformed interactive** group overall. Students preferred **interactive materials and geometric support**, with minimal reported cognitive overload. However, perceived understanding and preference for interactivity **did not improve actual performance**.

- Interactive design may have created **extraneous cognitive load**
- **Short learning phase**: students may have spent **too much time on manipulating simulations** and less on processing underlying concepts

Conceptual Understanding: No significant difference - understanding may have resulted from geometric support rather than from interactivity.
Problem-Solving: No significant difference - ceiling effect was present, due to student familiarity with the mathematical concepts and post-test questions not being challenging enough.
Knowledge Transfer: Static group performed significantly better - static materials may have encouraged deeper processing, while interactive feedback may have remained too transient for transfer learning.

7 Conclusion

Limitations:
- Small participant sample size
- Short learning phase time
- Ceiling effect in problem-solving section
- Late expert feedback results

Implications for Educators

- Interactive elements should not introduce extraneous cognitive load
- Pair visualizations with structured guidance: checkpoints, reflection questions, guided tasks
- Proportionally adjust the allocated learning time to account for interactive exploration

References:

[1] J. Shlens, "A Tutorial on Principal Component Analysis," arXiv, 2014.
[2] M. Greenacre, P. Groenen, T. Hastie, A. Iodice D'Enza, A. Markos, and E. Tuzhilina,"Principal component analysis," Nature Reviews Methods Primers, vol. 2, p. 100, 2022.
[3] Edoardo Saccenti. "A gentle introduction to principal component analysis using tea-pots, dinosaurs, and pizza", Teaching Statistics, 46:38–52, 2024.
[4] numiqo Team, "numiqo: Online Statistics Calculator," numiqo e.U., 2026. [Online]. Available: https://numiqo.com/lab/pca.
[5] V. Powell and L. Lehe, "Principal Component Analysis Explained Visually," setosa.io, 2015. [Online]. Available: https://setosa.io/ev/principal-component-analysis/
[6] S. Bechara, "PCA Builder & Visualization," MATLAB Central File Exchange, 2025. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/182689-pca-visualization-tool?s_tid=blogs_rc_6
[7] R. K. Atkinson, "Multimedia Learning of Mathematics," in The Cambridge Handbook of Multimedia Learning (R. Mayer, ed.), Cambridge Handbooks in Psychology, pp. 393–408,Cambridge: Cambridge University Press, 2005.
[8] 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, vol. 61, no. 9, pp. 2258–2290, 2024.

PCA Materials