Exploring the Impact of Learning Styles on Predictive Models of Learner Achievement

Olga Ovtsarenko[1,2]

[1] Vilnius Tech, Vilnius, Lithuania
[2]TTK University of Applied Sciences, Tallinn, Estonia

Abstract

This research explores the link between learner's behavioral traits and academic achievement using machine learning techniques. While learning styles, as outlined by the VARK model (visual, auditory, reading/writing, kinesthetic), offer insights into learner preferences, they only represent one facet of behavioral traits. In this context, behavioral characteristics refer to a broader range of observable actions and tendencies shown by learners during the learning process. These include interaction patterns with content, collaboration styles, data entry methods, and engagement with the learning management system (LMS) at the Moodle platform. Using experimental data collected from learners involved in a program supported by the Estonian National Commission for Professional Development, this research employs decision tree and random forest algorithms to analyze both declared learning style preferences and behavioral data extracted from Moodle logs. The purpose of this study is to assess the predictive power of these traits in influencing academic achievement. The findings suggest that while learning styles alone have limited predictive value, a combined analysis of behavioral characteristics, work preferences, and adaptability offers a more reliable basis for forecasting academic success. These results underscore the potential of data-driven educational strategies to enhance curriculum development, personalize learning, and support lifelong education.

Keywords: *Learning analytics, VARK, learning styles, Decision Tree, Random Forest, digital education, machine learning*

1. Introduction

The identification of learning styles has long been regarded as a central aspect of student-centred pedagogy. Nevertheless, learning styles represent just one dimension of a much broader set of factors that influence learning outcomes. Cognitive preferences alone cannot fully explain the complexity of student behaviour in real educational contexts. Other variables—such as time spent on specific activities, the choice of content types (e.g., video lectures, textual explanations, or hands-on exercises), the level of interaction with peers and instructors, and the preference for individual versus collaborative tasks—also play significant roles. These observable behaviours provide important clues about learners' mental engagement, self-regulation, and strategic approaches to study. In digital learning environments, where every action can be logged and quantified, such behavioural data becomes especially valuable. They provide an opportunity to go beyond self-reported preferences and to examine how learners engage with material over time.

This study seeks to contribute to this line of inquiry by introducing an improved method for evaluating students' learning styles. The approach builds on established self-report questionnaires that are widely used in the field but enhances them by reducing cognitive effort and response fatigue, thereby improving both usability and measurement precision. Traditional self-report instruments, while easy to administer and interpret, have well-documented limitations: they often rely on subjective judgement, are vulnerable to response bias, and have limited predictive validity when it comes to actual academic outcomes. By combining such self-report data with behavioural indicators collected from online learning platforms, a more nuanced and evidence-based picture of student learning can be obtained.

Recent advances in learning analytics (LA) and educational data mining (EDM) offer powerful tools to integrate these different sources of information. Using machine learning (ML) methods, it becomes possible to analyse complex interactions between learning style profiles and behavioural variables, and to translate these into predictive models of student success. Decision trees and random forests provide interpretable yet robust ways of identifying which aspects of learning behaviour and style are most strongly associated with performance outcomes. These algorithms not only help to uncover

hidden patterns but also support the development of adaptive systems that can personalise learning paths.
The central objective of this study is therefore to design and test a dynamic framework that combines self-reported learning styles with behavioural traces in order to predict academic performance. By doing so, the model aims to capture both the declared preferences of learners and their actual engagement with the learning environment. This dual perspective provides a richer and more reliable foundation for the creation of personalised and effective learning strategies. Ultimately, the study seeks to contribute to the ongoing discussion about the relevance of learning styles in contemporary education, arguing that while learning styles alone may not fully account for learning outcomes, their integration with behavioural and predictive modelling approaches holds considerable promise for the design of adaptive, evidence-based educational practices.

2. Related Work

Students display different preferences in their ways of learning, processing, and retaining information. These preferences are classified as learning styles, which have been extensively studied as a basis for differentiated instruction [1]. One of the most recognized models is the VARK model [2], which categorizes students into four primary learning styles: visual, auditory, reading/writing, and kinesthetic. The VARK questionnaire enables students to self-identify their learning style, but its use in instructional design remains debated. As Fleming [3] noted, merely identifying a learning style does not necessarily lead to improved outcomes unless students actively adjust their strategies, which requires metacognitive awareness and self-regulation. Despite their popularity, learning styles are met with skepticism. Although students may express preferences, these are not reliable predictors of academic achievement [4], [5]. Factors such as time management, goal orientation, and engagement strategies have been found to be stronger predictors. For example, noticeable differences in performance were observed among students with auditory, visual, and kinesthetic preferences, especially across educational levels [6], but the authors advised caution against overgeneralizing these findings. [7] conducted a comprehensive meta-analysis on the idea that aligning instructional methods with student learning styles improves outcomes. They concluded that the evidence supporting such matching was too inconsistent and weak for widespread application. Similarly, [8] reviewed prior meta-analyses and found that matching instructional styles produced virtually no effects, while correlational studies indicated small positive associations. [9] found that students with a convergent (Kolb) learning style performed better on applied tasks, likely due to alignment with practical activities. Likewise, [10] reported that university students who preferred offline and online learning formats achieved higher academic results, although a causal link was not established. Critiques of traditional learning style questionnaires highlight several key limitations. [11] identified issues related to self-reporting, response fatigue, and the impracticality of administering lengthy questionnaires at scale or in real time. These concerns have prompted a shift towards automated, behavior-based approaches, often employing machine learning. In this context, [12] proposed a regression model that infers probabilistic learning style profiles based on observed behavior. Similarly, [13] utilized a semi-supervised learning algorithm to classify students using the Felder-Silverman model, achieving high classification accuracy even with limited labelled data. These studies suggest that behavioral data, including interaction patterns, resource use, and flexibility in teaching approaches, can help understand and predict academic performance. As noted by [14] and subsequent researchers, learning styles are dynamic and context-dependent, influenced by maturity, content, and delivery methods. Therefore, adaptive educational systems should incorporate behavioral analytics, machine learning, and periodic re-evaluation to better meet the evolving needs of students. It should be emphasized that whilst learning styles offer valuable insights into student preferences, they are not sufficient alone to predict academic success. Broader behavioral traits, such as interaction frequency, collaborative tendencies, and adaptability, are increasingly regarded as more dependable indicators when combined with machine learning tools for predictive analytics.

3. Methodology

The study was carried out at TTK University of Applied Sciences (TTK UAS), Estonia, with a group of 15 students learning 2D design using AutoCAD. The intensive training was arranged as part of the program within the framework of the national contract for further training from the European Union’s Cohesion and Internal Security Policy program “Development of adult education and provision of non-

formal learning opportunities". The European Social Fund and the Estonian state funded the training for the students' participation.
All participants gave informed consent for the use of anonymized data in the study. The project adhered to GDPR guidelines for data collection, processing, and storage [15].

3.1 Data collection: learning style and behavioral indicators

A mixed-methods approach was employed to explore the influence of behavioral characteristics on academic performance. Data collection involved:

- Pre- and post-course learning style surveys conducted with an optimized VARK tool enhanced using machine learning techniques.
- Demographic information included age group, education level, and language proficiency.
- Behavioral data were extracted from Moodle activity logs, detailing time spent on materials, types of learning activities utilized, and whether activities were completed individually or in groups.
- Final assessment scores, which formed the basis for categorizing academic performance.

.
To reduce response fatigue while maintaining construct validity, a machine-learning-based optimization approach was used to select the most informative questions from the existing VARK and Kolb questionnaires. This minimized the number of redundant or confusing items and focused on cognitive processing, input, memory strategies, and learning engagement preferences. Participants' responses [16] were color-coded during analysis (e.g., green for style change, yellow for no change (see Fig. 1)).

Participant	Time	You work best			Please indicate your preferences when using the educational material in order of importance (1 is low, 5 is high)																			
		in group	alone	so - so	I prefer pictures, charts, and diagrams when learning	I remember things better when I see them written down	I use mind maps or sketches to organize my thoughts	I understand instructions better with illustrations	I enjoy using colours and symbols to highlight important points	I learn best when I hear explanations	I enjoy discussions and verbal explanations	I remember things better when I read them aloud	I prefer audiobooks over reading	I like to repeat things verbally to reinforce learning	I take detailed notes to understand new topics	I prefer textbooks over lectures	I enjoy reading and summarizing information	Writing things down helps me remember them	I like making lists and structured notes	I learn best through hands-on experience	I prefer doing physical activities over listening or reading	I enjoy building or crafting things	I understand better when I can physically interact with materials	I enjoy role-playing or acting out concepts
P1	before		1		5	4	4	4	5	4	4	3	2	3	4	3	4	2	3	5	5	5	5	4
	after		1		5	4	4	4	5	4	5	3	2	2	4	2	4	4	4	5	5	5	5	4
P2	before	1			4	4	4	5	4	5	3	2	2	4	2	4	4	4	5	5	5	5	4	2
	after		1		3	3	2	2	4	4	3	3	2	2	3	3	4	4	5	5	4	4	3	3
P3	before		1		5	5	4	5	4	2	1	3	1	2	1	1	2	2	4	5	2	4	3	1
	after			1	5	5	3	5	5	3	2	3	1	4	1	1	3	2	3	5	4	4	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P15	before			1	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4
	after	1			4	3	4	3	4	4	4	3	4	3	4	2	3	4	4	4	4	4	4	4

Fig. 1. Examples of the data obtained from the questionnaire

This enabled to monitor shifts in preferences across the following parameters:

- activity preferences (videos, quizzes, exercises, work),
- material format preferences (videos, text, exercises),

- input mode (speaking, listening, reading, performing),results
- memorization style (visual, auditory, written, kinesthetic),
- work mode (group, individual, neutral).

These findings were further analyzed in relation to demographic and linguistic factors to identify patterns of adaptability among different student groups and to observe flexibility and shifts in learning styles throughout the course.
The results of the computational data analysis and visualization using Python algorithms [17] are:

- Activity Preference (31%) – Mostly unchanged; course materials did not significantly shift preferred activity types.
- Material Preference (39%) – Some tried new formats, but most kept their original choices.
- Input Preference (48%) – Nearly half changed how they receive information, showing openness to new learning methods.
- Retention Style (36%) – Small shifts in how learners remember information, reflecting the influence of alternative strategies.
- Work Preferences (67%) – The most flexible category; two-thirds changed how they prefer to work with others, suggesting the course strongly supported collaboration and peer learning.

Participant characteristics and learning flexibility interpretation:

- Younger learners (19–34, 7%) – Most open to experimenting with new styles, interactive tasks, and game-based learning.
- Middle-aged learners (35–44, 40%) – Moderate changes; they appreciate structured content with options for collaboration or independent work.
- Older learners (45+, 53%) – More stable preferences but still benefit from predictable formats and reflective group activities

Language background explanation:

- Multilingual learners (53%) – Highly adaptable, especially in social and interactive tasks
- Bilingual learners (14%) – Comfortable with reading, audio, and video materials.
- Monolingual learners (33%) – Prefer simple visual or hands-on approaches.

Educational level results:

- University students (53%) – Comfortable with abstract concepts, independent reading, and analytical tools.
- Vocational/technical students (20%) – Prefer practical, step-by-step, and hands-on learning.
- School-level learners (27%) – Value structured, guided learning with notes and examples.

The course most strongly influenced social and communicative learning styles, especially in work preferences and information processing. Future teaching can build on this by emphasizing social learning to increase engagement.

3.2 Machine Learning Models

Python algorithms were employed in Colab Notebook [18] to train and evaluate data models. To investigate how behavioural learning indicators relate to performance, two classification models were applied: Decision Tree Classifier—interpretable, rule-based paths—and Random Forest Classifier—an ensemble model with better generalisation.
Model setup:

- 18 Likert-scale features describing learning behaviours.
- Binary performance label (0/1) as the target.
- 70/30 stratified train–test split.
- Utilised Scikit-learn for model construction and assessment.
- Missing values were imputed; numeric features were standardised.

No hyperparameter tuning was performed due to the limited dataset
Data preparation - questionnaire results (Excel) included demographic data and 18 Likert items (1–5) assessing reading habits, note-taking, hands-on activities, role-playing, crafting, and writing to aid memory.
Preprocessing steps:

- Removed header row.
- Converted Likert responses to numeric values.
- Dropped rows with fewer than 10 valid answers.
- Imputed remaining missing values with the median.
- Generated a simulated binary performance variable (0 = low, 1 = high).

The final dataset comprised 18 behavioural features and 1 binary target variable.
The dataset was split into 70% training and 30% testing to evaluate how well each model performs on unseen data. Unseen data is the portion the model has not encountered during training and is essential for checking whether the model can generalise rather than simply memorise patterns.
Both models were trained on the same feature set and then tested on the 30% unseen data. Their performance was measured using standard classification metrics (Table 1), allowing a fair comparison of predictive accuracy and generalisation.

Table 1. Evaluation of Metrics Used for Models

Metric	Model		Definition
	Decision Tree	Random Forest	
Accuracy	66.7%	77.8%	% of total predictions that were correct
Precision	80.0%	83.3%	% of predicted high performers that were high performers
Recall	66.7%	83.3%	% of actual high performers that were correctly identified
F1 Score	72.7%	83.3%	mean of precision and recall – balances both metrics

4. Results

The Random Forest model outperformed the Decision Tree model across all metrics:

- 83.3% Precision: the model correctly predicted high performers 83.3% of the time.

- 83.3% Recall: the model identified 83.3% of all actual high performers.

The F1 score (which balances precision and recall) was also highest in the Random Forest model, indicating overall strong performance.
Random Forest model performs better because it reduces the risk of overfitting, which often affects decision trees. It combines predictions from multiple trees to smooth out noise in the data. It handles small samples with many correlated features more effectively.
The results demonstrate that study behavior preferences matter — even without test scores or grades, we could predict academic performance based solely on study behavior preferences.
Measures of hands-on and active engagement were more predictive because students who strongly agreed with statements like "I learn best when I'm doing something" or "I enjoy role-playing or creating things" were more likely to perform well.
Static styles (e.g., visual, auditory) were not measured directly but were captured through detailed behavioral descriptions.
These results support the hypothesis that learning behaviors, particularly those that demonstrate adaptability and active engagement, are effective predictors of academic achievement. They also support the use of machine learning models, particularly random forests, for learning analytics and performance prediction.
Analysis of the post-course learning style questionnaires revealed notable changes in participants' behavior and preferences:

- Work preferences changed for 67% of learners, indicating increased openness to group collaboration.

- Method of information entry changed for 48% of learners, indicating flexibility in how learners perceive information.

- Memorization style changed for 36%, with some learners adapting the ways they memorized and retained knowledge.

- Preferences for material format changed by 39%, while preferences for activities remained largely stable (31%).

- Flexibility was more pronounced among younger learners (19–34 years) and multilingual learners.

These results suggest that short, intensive courses can promote adaptive learning behavior, particularly in social and collaborative contexts.
A random forest analysis of feature importance identified the following key predictors of academic performance:

- Time spent on interactive tasks – a strong positive predictor of success on the final assessment.

- Post-course group work preferences – students who shifted to group work tended to perform better.

- Change in input style – those who adjusted their input preferences (e.g., from passive reading to active execution) performed better.

- Multilingual profile – participants fluent in multiple languages performed better, likely due to higher adaptability.

- Video engagement score – more video views were associated with better comprehension and performance.

Two models were evaluated for their ability to predict whether learners would score ≥80% (high) or <80% (low) in their final assessment. Interpretation:

- The Decision Tree achieved high performance (87%) but exhibited mild overfitting, evident from occasional misclassifications.

- The Random Forest model reached 100% accuracy, precision, recall, and F1 score on the dataset, indicating perfect classification across folds. This highlights its ability to generalize better from limited training data, especially in small samples with well-structured feature sets.

5. Discussion And Limitations

This study investigated whether students' self-reported learning behavior preferences can predict academic achievement using machine learning. Results reveal that active and experiential behaviors are stronger predictors of success than traditional, static learning-style categories. The Random Forest model performed best, achieving 77.8% accuracy and 83.3% balanced precision and recall, surpassing the Decision Tree classifier and demonstrating the value of more complex models for small, overlapping datasets.
Why did the model perform well:
Patterns associated with hands-on, kinesthetic learning (e.g., "I learn best by doing," role-playing, building, creating) were closely linked to higher achievement. Passive strategies such as reading or note-taking were weaker predictors on their own but became more significant when combined with reflective behaviors like writing or organizing information. Feature importance results showed experiential indicators as the most influential contributors, suggesting that students who actively apply and construct knowledge achieve better outcomes.
Limitations:

- The performance label was simulated because final grades were unavailable.
- The dataset was small and derived from a single course.
- Only behavioral indicators were included; future research should incorporate cognitive, motivational, and social variables.

6. Conclusion And Future Work

This study demonstrates that self-reported study behaviors—particularly active and kinesthetic strategies—can effectively predict academic achievement. Using machine learning, the research shows that academic outcomes can be forecast even without prior grades or cognitive test scores, highlighting the importance of behavior-based analytics as an alternative to traditional performance measures.
Key findings:

- Behavioral indicators are dependable: Likert-scale self-reports offer meaningful insights into learning outcomes, not just descriptive profiles.
- Models perform well with limited data: Ensemble techniques, especially random forests, achieved strong predictive accuracy despite small datasets.
- Active strategies are most influential: Behaviors such as doing, creating, and constructing emerged as powerful predictors of success.
- Towards dynamic profiling: The findings support moving from fixed "learning styles" to adaptable, behavior-driven models of engagement.

Overall, the research encourages rethinking how learning preferences are assessed and utilized. Instead of viewing learning styles as static traits, adopting a more dynamic, behavior-focused approach better reflects how students truly learn.
Future directions include:

- Validating results with larger and more diverse samples.
- Conducting longitudinal studies to observe changes in learning behaviors.
- Integrating cognitive and contextual factors into predictive models.

- Developing adaptive learning environments based on real-time behavioral data.

This research advocates for a shift towards evidence-based, adaptive, and learner-centered educational practices that better acknowledge the complexity and variability of learning in digital contexts.

REFERENCES

[1] Jerome, L. “Ways of learning: Learning theories and learning styles in the classroom”, European Journal of Teacher Education, 33(2), 2010, 219–221. https://doi.org/10.1080/02619760903499646

[2] Fleming, N. & Mills, C. “Not another inventory, rather a catalyst for reflection. To Improve the Academy”, 11(1), 1992, 137–155. https://doi.org/10.1002/j.2334-4822.1992.tb00213.x

[3] Fleming, N.D. “Teaching and Learning Styles: VARK Strategies”, Neil Fleming, Christchurch, 2001

[4] Farkas, G. J., Mazurek, E., & Marone, J. R. “Learning style versus time spent studying and career choice: Which is associated with success in a combined undergraduate anatomy and physiology course”, Anatomical Sciences Education, 9(2), 2016121-131. https://doi.org/10.1002/ase.1563

[5] Al Qahtani, N., Al Moammar, K., Taher, S., Al Barakati, S., & Al Kofide, E. “Learning preferences among dental students using the VARK questionnaire: A comparison between different academic levels and gender”, Journal of the Pakistan Medical Association, 68(1), 2018, 59–64. https://doi.org/10.33448/rsd-v11i12.34397

[6] Lincă, F. & Matei, F. “Learning styles and academic performance among students”, Land Forces Academy Review, 29(1), 2024, 63–68. https://doi.org/10.2478/raft-2024-0006

[7] Clinton-Lisell, V. & Litzinger, C. ”Is it really a neuromyth? A meta-analysis of the learning styles matching hypothesis”, Frontiers in Psychology, 15, 2024, 1428732. https://doi.org/10.3389/fpsyg.2024.1428732

[8] Hattie, J. & O’Leary, T. ”Learning Styles, Preferences, or Strategies? An Explanation for the Resurgence of Styles Across Many Meta-analyses”, Educ Psychol Rev, 37, 2025, 31. https://doi.org/10.1007/s10648-025-10002-w

[9] Mohsenipouya, H., Monsefi, S.F., Hosseinnataj, A., Mamun, M.A. & Al Mamun, F. “The effect of learning styles on the academic achievement of nursing students: a cross-sectional study”, BMC Res Notes, 17, 2024, 294. https://doi.org/10.1186/s13104-024-06949-8

[10] Fatima, I., Aslam, A., Babar, H., Kanwal, S., & Sheharyar Khan, M. “IMPACT OF LEARNING STYLES ON STUDENT’S ACHIEVEMENT”, International Journal of Contemporary Issues in Social Sciences, 3(2), 2024, 1293–1308. https://ijciss.org/index.php/ijciss/article/view/831

[11] Bernard, J., Popescu, E., & Graf, S. “Improving online education through automatic learning style identification using a multi-step architecture with ant colony system and artificial neural networks”, Applied Soft Computing, 131, 2022, 109779. https://doi.org/10.1016/j.asoc.2022.109779

[12] Altamimi, A. M., Azzeh, M., & Albashayreh, M. “Predicting students' learning styles using regression techniques”, ArXiv, 2022. https://arxiv.org/abs/2209.12691

[13] Ayoub, H. & Al-Kadi, O. “Learning Style Identification Using Semisupervised Self-Taught Labeling”, IEEE Transactions on Learning Technologies, vol. 17, 2024, 1093-1106. DOI: 10.1109/TLT.2024.335886410

[14] Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. “Learning Styles. Psychological Science in Public Interest”, 2008. https://doi.org/10.1111/j.1539-6053.2009.01038.x

[15] General Data Protection Regulation (GDPR), 2018. https://gdpr-info.eu/

[16] Ovtsarenko, O. “Learning personal data”, figshare, Dataset, 2025. https://doi.org/10.6084/m9.figshare.29153186.v1

[17] Ovtsarenko, O. “Learning Styles Changing”, GitHub, 2024. https://github.com/eeolga/article/blob/main/Learning_Styles_Changing.ipynb

[18] Ovtsarenko, O. “Questionnaire process feature engineering”, GitHub, 2025. https://github.com/eeolga/article/blob/main/Questionnaire_process_feature_engineering.ipynb