ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #45369

# The Impact of AI Assistance on Student Learning: A Cross-Disciplinary Study in STEM Education

**Prof. Matthew Fried, SUNY Farmingdale**

Matthew Fried is an Assistant Professor with a research focus in machine learning. His work includes the application of advanced mathematical techniques, such as the Choquet integral, to deep neural networks (DNNs). He has presented multiple papers on this topic at international conferences, contributing to the ongoing development of noise reduction and performance optimization in DNNs.

# The Impact of AI Assistance on Student Learning: A Cross-Disciplinary Study in STEM Education


## Abstract

We examine the widespread use of AI in the form of Large Language Models (LLMs) as a tool for academic assistance. The study investigates whether students studying with AI assistance retain more information compared to those employing standard alternative approaches such as using a basic search engine, reviewing with a friend, or contemplating the material independently. The research reveals that while basic tasks and retention may benefit from AI assistance, outsized gains are lacking. Counterintuitively, specific tasks related to deep thinking and conceptual exploration are found to be better served with alternative approaches. We compare different features via hypothesis testing (p-values), ANOVA, logistic regression, and chi-square, highlighting relationships and the lack thereof. The data was collected across multiple colleges in various STEM disciplines, providing a robust cross-disciplinary perspective. Additionally, the paper discusses the influence of ethnic and cultural background, learning styles, technical talent, and other contributing factors to student success when utilizing LLMs.


## Introduction

Advances in technology have introduced a range of new learning methods, with AI-based platforms gaining increasing attention in education; and while peer-based learning and internet-based learning are also popular among students seeking flexible, interactive ways to enhance their understanding, we seek to explore the efficacy of the latest trend. Thus, the question remains: which learning approach is most effective in improving learning outcomes?

This study compares several distinct learning approaches, highlighting the following: reviewing with AI, reviewing with internet-based assistance, and reviewing with a friend. We also include personal contemplation and a free choice as well, but mostly categorize it as *other* in our evaluation of impact on student performance. By using predictive models and non-parametric tests, we aim to identify which factors significantly contribute to the effectiveness of each learning approach and which method produces the best results overall.

This study explores the effectiveness of various learning approaches using a dataset that includes key factors such as ancestry, learning process, and technical talent. We applied non-parametric tests (Kruskal-Wallis) and logistic regression models with L1, L2, and Elastic Net regularization to compare the effectiveness of each approach. Results suggest that learning processes such as reading/writing and visual learning significantly influence the success of AI-based learning, while other methods show a strong association with early adapter tendencies among other statistically significant results. The Bonferroni correction was applied to ensure that the multiple comparisons conducted remained statistically valid, with all results to follow.

**Literature Review**

The rise of AI in education has been extensively researched, with studies focusing on AI's role in providing personalized feedback, enhancing student engagement, and offering scalable learning solutions. As such, the integration thereof has emerged as a transformative force, reshaping pedagogical approaches and learning experiences. Various studies highlight the significance of AI literacy, student readiness, and the adaptation of educational frameworks to accommodate the above based on diverse learning styles, backgrounds, and more.

Ng et al. [1] emphasize the necessity of defining AI literacy, which encompasses understanding AI's capabilities and ethical implications, as well as evaluating its impact on learning environments. The study proposed four aspects: know and understand, use, evaluate, and ethical issues, for fostering AI literacy based on the adaptation of classic literacies. We use similar competency targets in our study as well, focusing specifically on learning. Furthermore, Dai et al. [2] discusses student confidence in learning AI concepts and recognizing the relevance of AI knowledge in their lives, with results indicating that AI literacy was not predictive of AI readiness, which in this paper we explore in terms of actual implementation at a practical learning level of AI usage versus standard methods. Beyond this, Lim [9] underscores the benefits of fostering a sense of ownership over student educational experiences, with autonomy playing a crucial role for motivation and engagement.

Chiu and Chai [3] discuss the importance of curriculum planning for AI education, highlighting the need for teachers to be equipped with appropriate resources and strategies for engagement. They used self-determination theory (SDT), with the four basic structures of content, product, process and praxis to reveal that genuine curriculum creation should encompass all four forms of design, to be orchestrated around student learning experiences. Thus, in this paper, we question whether students will effectively build such skills on their own, using an evaluative questionnaire meant to similarly evoke self-determination theory structures, recognizing that student implementation may be more naturally adaptive and therefore better [4].

Katsaris and Vidakis [5] emphasize the importance and efficiency of the utilization of learning styles in the adaptive learning process, discussing how to help students discover for themselves what environments they can learn best in and how machine learning techniques can assist them to efficiently extract, analyze and adapt in education and other venues, with personalization a key component to learning. Others such as Al Hadithy at al. [6] note that students wish to have AI literacy taught in their programs, specifically that there is demand for AI to be included in medical education.

In terms of learning, Alshammari et al. [7] highlight that a high level of adaptation in e-learning systems enhances usability, which can lead to increased learner satisfaction and engagement, critically so. They state, “that an adaptive e-learning system based on learner knowledge and learning style has a higher level of perceived usability than a non-adaptive e-learning system.” And they go on to say that this may also increase the level of satisfaction, engagement and motivation. Extending this with much success, Xu [8] created a personalization scheme as an adaptive learning system for vocabulary acquisition that utilizes machine learning to tailor content to individual learners, with the results showing that “the proposed learning system outperforms previous approaches in terms of learning efficiency, scientifically, and reliability.”

In contradistinction to the above, social constructivism posits that knowledge is constructed through social interactions and shared experiences, mirroring the general peer-based learning models that facilitate deeper understanding through said perspectives, with researchers such as Zhai [10] highlighting the benefits of social interaction in learning and mapping characteristics to machine learning where they conclude that machine learning has “transformed—but not yet redefined—conventional science assessment practice in terms of fundamental purpose, the nature of the science assessment, and the relevant assessment challenges,” a challenge to the current educational system to integrate such tools and still foster peer-based learning. Seo [11] also mentions that “there were concerns about responsibility, agency, and surveillance issues,” while nevertheless encouraging AI use for many applications. Other researchers, such as Elwarraki [12] discuss teachers as facilitators in these personalized environments in the general education system, bringing to light models that may find collaborative opportunities and experiences, as in Kim [13].

Thus, in summary, we note the benefits of adaptive learning systems and the potential ability to effectively leverage machine learning algorithms to tailor content to individual students’ needs, resulting in better outcomes for students with diverse learning styles as compared to peer-based learning, rooted in social constructivism, emphasizing collaboration and mutual knowledge-building among students, which perhaps is not yet a strength of AI assisted study. Furthermore, as a comparison, we study internet-based learning which provides the flexibility of self-directed learning, allowing students to access a wide range of resources at their own pace, seemingly a different modality then AI study, as explained above.

This paper contributes to the existing body of work by conducting a comparative analysis of these learning approaches, incorporating statistical methods to identify which variables are most predictive of success in each approach.

**Methodology**

**3.1 Data**

The dataset used in this study includes several variables that influence learning outcomes, including:

- Hot_Or_Cold: How is a student currently feeling. Range: 1-5. Used as a test to gauge test conditions comfort level and potential outlier activity.
- Ancestry: With options of Western European, Black or African, Hispanic or Latino, Middle Eastern or North African, Asian, Native American, Slavic, or I prefer not to say.
- Technical_Talent: Assessed technical skills. Range: Terrible (1) to Wonderful (5).
- Learning_Process: Learning style. Range: Visual, Auditory, Reading/Writing, or Kinesthetic.
- Learning_Approach: Learning method. Range: Collaborative, Experiential, or Observation.
- Early_Adapter: Whether the student tends to adopt new technologies early. Range: Yes or No.

The target variable, Test_Group, categorizes students into those who review with AI, review with the internet, review with a peer, contemplate on their own, or choose any method they like.

We then asked a set of 10 to 12 questions (usually 12) broken into four categories: knowledge recall, procedural knowledge, application and problem solving, and analysis and evaluation. The two former focused on technical understanding and skill and the two later focused on the ability to draw conclusions through analysis. Each score was normalized (with a minimum of 0 and a maximum of 1) and then added together to create a total score with a range from 0 to 4.

The study included 224 students across multiple schools (which will remain anonymous to protect the students' privacy). Students were given the questionnaire in an Excel course, a web programming course, a data science course, a quantum computing course, and a course introducing the Microsoft suite. Students were taught the material for the day, were asked to study for five minutes as one of the Test_Group categories, and then were given the questionnaire.

**3.2 Statistical Tests**

Despite having 224 participants, within several tests the data was imbalanced and therefore techniques such as SMOTE (Synthetic Minority Over-sampling Technique) were applied to augment the dataset, avoid biased models, fix against poor recall, reduce overfitting, and ensure sufficient instances to study of minority cases.

Initially, the data was studied in Excel and then moved into a Python environment to do Chi-Square tests to identify significant relationships with p-values below a 5% threshold.

Several models were tested to target the Test_Group variable, such as Random Forest Classifier, XGBoost, and a simple Neural Network. These models were quite weak, therefore Test_Group was broken into several sets, (1) Review with Internet vs Other, (2) Review with AI vs Other, (3) Review with Friend vs Other, in order to better understand the data via Logistic Regression and use L1 Regularization (Lasso), used to select the most important features by shrinking irrelevant coefficients to zero, L2 Regularization (Ridge), applied to reduce multicollinearity and improve the stability of the model, and Elastic Net, a combination of L1 and L2 regularization, offering a balance between feature selection and regularization.

We note that ANOVA was also used as a parametric test against the categories of learning vs Test_Group, as well as the Kruskal-Wallis Test, a non-parametric test to determine if there are statistically significant differences between the groups (using the categorical data). We applied this test to compare learning outcomes for AI, peer-based review, and internet-based review, particularly focusing on variables like Ancestry and Learning Process.

Bonferroni Correction was applied as well, to control for Type I errors, by dividing the significance threshold by the number of comparisons. We did this to ensure that the likelihood of false positives was minimized.

## Results

### 4.1 Initial Exploration

Our initial results show that the average review score for students reviewing with AI was higher than other approaches with the score having a more than one standard deviation above the mean, significantly higher than everything but reviewing with the internet. Furthermore, we note that while reviewing with AI and the internet did well on knowledge recall and procedural knowledge, contemplating by oneself did best in analysis and evaluation.

We also note that those students that rated themselves with higher technical talent did better at knowledge recall while those that rated themselves as having worse technical talent did better at problem solving tasks. Additionally, those with weak technical talent also did better with AI review than those with stronger technical talent.

In terms of learning approaches, experiential learners did best with AI review while collaborative learners did best with internet review.

While these initial results are not necessarily statistically significant, we mention them only to give a broad understanding of the data and to inspire further research as well. Now, as to our statistical analysis, we begin by listing those variables that has a significant p-value against the test group (see Figure 1):

| Chi-Square Test for each variable agasinst Test_Group | |
|---|---|
| Ancestry | Chi-squared test statistic: 41.8457<br>- Degrees of freedom: 24<br>- P-value: 0.0134 |
| Technical Talent | - Chi-squared test statistic: 32.2522<br>- Degrees of freedom: 16<br>- P-value: 0.0093 |
| Application and Problem Solving | - Chi-squared test statistic: 1.8267<br>- Degrees of freedom: 4<br>- P-value: 0.7676 |
| Analysis and Evaluation | - Chi-squared test statistic: 0.8045<br>- Degrees of freedom: 4<br>- P-value: 0.9378 |

*Figure 1 – Chi-Square Tests*

We then did one-hot encoding on our categorical data with a multinomial logistic regression on Test_Group and found that the 3 significant factors here were Learning_Approach_Experiential learning, Early_Adapter_Yes, and Technical_Talent_4.0, each with a p-value below 5%.

Below (see Figure 2) is a summary chart of the statistics of testing AI vs Other, Internet vs Other, and Friend vs Other. Each model was reasonable, with a definite room for improvement. Note the strongest features with the log-odds for each category. For the first two models, predictors included Learning_Approach and Learning_Process, while for the third model, Ancestry and Early_Adapter_Yes were the most prominent.

After this, we tested L1, L2, and Elastic Net in the Logistic Regression (with optimized hyperparameters) and we were able to improve our models. The optimal setting for:

- AI vs Other was an L1 regression that gave an accuracy of .69,
- Internet vs Other was an L2 regression that gave an accuracy of .62,
- Friend vs Other was an L2 regression that gave an accuracy of .78.

| Logistic Regression | | | |
|---|---|---|---|
| AI vs Others | | | |
| | precision | recall | f1-score |
| 0 | 0.63 | 0.67 | 0.65 |
| 1 | 0.68 | 0.64 | 0.66 |
| accuracy | 0.66 | | |

| Features | Log Odds |
|---|---|
| Learning_Approach_Observation and Imitation | 1.333 |
| Learning_Approach_Experiential learning | 1.242 |
| Ancestry_Western European | 1.137 |
| Ancestry_Slavic | 0.914 |

| Internet vs Others | | | |
|---|---|---|---|
| | precision | recall | f1-score |
| 0 | 0.76 | 0.44 | 0.56 |
| 1 | 0.51 | 0.81 | 0.63 |
| accuracy | 0.6 | | |

| Features | Log Odds |
|---|---|
| Learning_Process_Reading/Writing | 1.509 |
| Learning_Approach_Experiential learning | 0.87 |
| Learning_Process_Visual | 0.834 |

| Friend vs Others | | | |
|---|---|---|---|
| | precision | recall | f1-score |
| 0 | 0.8 | 0.73 | 0.77 |
| 1 | 0.7 | 0.77 | 0.73 |
| accuracy | 0.75 | | |

| Features | Log Odds |
|---|---|
| Early_Adapter_Yes | 1.874 |
| Ancestry_Middle Eastern or North African | 1.45 |
| Learning_Process_Visual | 0.1348 |

*Figure 2 - Logistic Regression Models*

Our next results were for the p-values of less than 5% for each of the three variations tested above. We found that ancestry played a significant role in different models, perhaps as an outgrowth of culture norms or the like.

| AI vs Other | P-Value |
|---|---|
| Ancestry_Western European | 0.006 |
| Learning_Process_Reading/Writing | 0.029 |
| Learning_Process_Visual | 0.017 |

| Internet vs Other | P-Value |
|---|---|
| Ancestry_Hispanic or Latino | 0 |
| Ancestry_Slavic | 0.009 |
| Learning_Process_Reading/Writing | 0.015 |

| Friend vs Other | P-Value |
|---|---|
| Ancestry_Western European | 0.028 |
| Early_Adapter_Yes | 0 |

*Figure 3 – P-Values for Comparison Models*

Further test bore out the two factors (ancestry and learning process) as the critical factors in many of the reviewing aspects, with the Kruskal-Wallis test giving similar results, wherein p-values for AI vs Other were significant for ancestry and learning process; and wherein p-values for Friend vs Other were significant for Early_Adapter.

And as a final result, fixing against false positives but possibly drowning out the information in the data, we used the ultra-conservative Bonferroni adjustment to the Kruskal-Wallis test and still found that Learning_Process was a significant factor in AI vs Other.

**Discussion of Results**

The findings from our study indicate that learning approaches, particularly AI-based learning, are strongly influenced by specific factors such as Ancestry and Learning Process. Students from Western European backgrounds, for instance, may find AI platforms more conducive to learning, while peer-based review is more effective for students identified as Early Adapters.

The combination of non-parametric tests (Kruskal-Wallis) and regularized logistic regression models allowed us to identify significant predictors while controlling for false positives through the Bonferroni correction. This has shown that students using AI that have a learning approach related to observation and imitation as well as experiential learning will do well with AI, while students more geared towards reading and writing might do better studying with the internet, while yet early adapters of technology as well as visual learners would do best by studying with a friend.

In general, as per the output of several models, experiential learners have an advantage when it comes to AI and technology, while students in general need to be careful when using an AI, as they may gain in knowledge recall and procedural knowledge but would be better off thinking on their own to advance their skillset in analysis and evaluation.

**Conclusion**

Our study highlights the importance of personalizing learning approaches based on key factors such as Ancestry, Technical Talent, and Learning Process. While AI-based learning shows promise for certain student groups, peer and internet-based reviews also play a vital role in fostering engagement and knowledge retention. To this end, students should be wary of entirely relying on AI, as backgrounds, learning preferences, and deep analysis may be hurdles instead of older, standard approaches. Future research should explore the interactions between these variables in greater detail, perhaps using larger datasets and different learning environments.

**References**


[1] Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). ai literacy: definition, teaching, evaluation and ethical issues. Proceedings of the Association for Information Science and Technology, 58(1), 504-509. https://doi.org/10.1002/pra2.487

[2] Dai, Y., Chai, C. S., Lin, P., Jong, M. S., Guo, Y., & Jian-jun, Q. (2020). Promoting students' well-being by developing their readiness for the artificial intelligence age. Sustainability, 12(16), 6597. https://doi.org/10.3390/su12166597

[3] Chiu, T. K. F. and Chai, C. S. (2020). Sustainable curriculum planning for artificial intelligence education: a self-determination theory perspective. Sustainability, 12(14), 5568. https://doi.org/10.3390/su12145568

[4] Cavanagh, T. B., Chen, B., Lahcen, R. A. M., & Paradiso, J. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: a practitioner's perspective. The International Review of Research in Open and Distributed Learning, 21(1), 172-196. https://doi.org/10.19173/irrodl.v21i1.4557

[5] Katsaris, I. and Vidakis, N. (2021). Adaptive e-learning systems through learning styles: a review of the literature. Advances in Mobile Learning Educational Research, 1(2), 124-145. https://doi.org/10.25082/amler.2021.02.007

[6] Al Hadithy, Z. A., Al Lawati, A., Al-Zadjali, R., & Al Sinawi, H. (2023). Knowledge, attitudes, and perceptions of artificial intelligence in healthcare among medical students at sultan qaboos university. Cureus. https://doi.org/10.7759/cureus.44887

[7] Alshammari, M. T., Anane, R., & Hendley, R. J. (2015). Design and usability evaluation of adaptive e-learning systems based on learner knowledge and learning style. Lecture Notes in Computer Science, 584-591. https://doi.org/10.1007/978-3-319-22668-2_45

[8] Xu, Y. (2022). An adaptive learning system for english vocabulary using machine learning. Mobile Information Systems, 2022, 1-9. https://doi.org/10.1155/2022/3501494

[9] Lim, S. H., Lim, L., & Lim, R. W. Y. (2023). Learners' satisfaction with adaptive learning systems. Educational Innovations and Emerging Technologies, 3(2). https://doi.org/10.35745/eiet2023v03.02.0001

[10] Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., & Urban-Lurain, M. (2020). From substitution to redefinition: a framework of machine learning-based science assessment. Journal of Research in Science Teaching, 57(9), 1430-1459. https://doi.org/10.1002/tea.21658

[11] Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. International Journal of Educational Technology in Higher Education, 18(1). https://doi.org/10.1186/s41239-021-00292-9

[12] Elwarraki, O., Aammou, S., & Lahiassi, J. (2023). Teachers as facilitators. Conhecimento &Amp; Diversidade, 15(39), 146-155. https://doi.org/10.18316/rcd.v15i39.11161

[13] Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-ai collaboration: perspectives of leading teachers for ai in education. Education and Information Technologies, 27(5), 6069-6104. https://doi.org/10.1007/s10639-021-10831-6