



From Prompt Revision to Learning Gains: Modeling Self-Regulated AI-Supported Learning Using a Machine Learning Algorithm

¹Elugbadebo O. Joseph, ²Femi T. Johnson, ¹Sunday A. Akinyele, ¹Akiode Janet

¹Department of Computer Science Education, Federal College of Education, Abeokuta, Ogun.

²Department of Information Technology and Health Informatics, Federal University of Medicine and Medical Sciences, Abeokuta, Ogun

Corresponding Author: femi.johnson@fummsa.edu.ng

Abstract

Self-regulated learning behaviours among students in the digital age measures how well they revise their prompts. This study uncovers the correlation of prompt revision behaviours to learning gain among 100 undergraduates who had a 2-week structured ChatGPT course. Revision behaviours were classified into six groups (no revision, shallow, constraint addition, format change, decomposition, example-driven). A Self Regulation Score (SRS) from revision depth and persistence was constructed for correlation between pre and post test gain. A machine learning algorithm (Random Forest) was utilized to predict learning gains (top range) from revision features. Results obtained reveals SRS (n=100, and 1,620 interactions) correlated strongly with learning gain ($r=.87$, $p<.001$). Random Forest achieved Accuracy and Recall scores of 75.6%, Precision of 80%, and F1-score 73%. The findings from this study support the development of diagnostic system to aid educators identify prompt technique needs of students and integrate real-time self-regulation analytics into generative AI classrooms.

Keywords: Self-regulated learning, prompt revision, ChatGPT, Orange Data Mining, learning analytics

Introduction

The inclusion and utilization of digital and information communication gadgets in schools has brought a significant improvement in the mode of teaching and learning (Elugbadebo & Johnson, 2020). Educational organizations including primary, secondary and tertiary institutions are investing greatly to digitalize educational processes with supports from individuals, corporate organizations and the government. The outcome of earlier investment on digital and ICT tools have led to improved student performance, and reduction in manual work load. As these improvements become noticeable and evidenced, a new trend is the utilization of Generative Artificial Intelligence (GenAI) in education (Lo, 2023). Researchers and authors have posed different views on the acceptability (Zhao, 2022) and utilization of GenAI tools especially in higher education (Dwivedi et al., 2023).

Students' utilization of large language tools and GenAI including ChatGPT transcends through numerous forms (Tlili et al., 2023) from image and text generation, concepts explanation, critiques

arguments to code generation and debugging (Chen et al., 2024). However, the quality of the results derived is dependent on the quality of user's prompt. When GenAI responses mismatch students' knowledge and have little or no significant relevance with students' prompts, it may lead to frustration and abandonment of the task (Carrasco-Sáez et al., 2025). Although, some students respond by strategically changing roles, revising prompts, simplifying questions or adding more information for clarity to the GenAI yet many students lack the capacity to recognize poor outcomes and adjust behaviour which formed the hallmark of self-regulated learning (Zimmerman, 2002).

Panadero, (2022) described self-regulated learning (SRL) as an intentional change in knowledge acquisition, based on experience or practical experiment. It is typically measured via surveys or think-alouds, which are labour-intensive and not scalable. This form of regulated learning imbibes assertive and proactive skills needed to control and achieve learning goals among students.

Recent work suggests that young learners' log-based behavioural indicators of revision patterns can capture SRL in digital environments (Molenaar, 2022). Self-regulated learners complete a cyclic phase requiring one or more processes involving planning, monitoring, control and evaluation (Viberg et al., 2023). This can be achieved when students simplify large complex questionnaires or text before prompting on ChatGPT (Farrokhnia et al., 2024). Some students develop or propose evidence-based responses while adjusting prompts appropriately. Prompt revision depth after a poor AI response correlates with quality of response. Chang (2023) discovered that students who perform socratic prompting techniques and include constraints achieve higher task completion rates with improved conceptual understanding. However, no study has linked prompt revision behaviour directly to learning gains measured by objective pre-/post-tests, nor provided a transparent, no-code toolkit for classroom use. This study addresses these two research questions below with a 2-week undergraduate course with four ChatGPT exercises collected through interaction logs, coded revision behaviour using a simple and modified rule based taxonomy, computed SRS in Orange Data Mining software, and evaluated predictive performance.

1. To what extent does Self-Regulation Score (SRS) derived from prompt revision behaviour correlate with learning gain over a 2-week course?
2. Can a model using only revision features (no prior grades) predict which students will achieve high learning gain?

Literature Review

Zimmerman (2002) described a self-regulated learner as one involved in planning, monitoring, and adjusting his cognitive strategies for academic achievement. The present day technology-driven environments allow students to learn using the SRL techniques. Common techniques utilized range from pausing, re-reading, help-seeking to revision patterns (Molenaar, 2022). With generative AI, a new SRL behaviour emerges known as prompt revision. Through this prompt revision, a self-regulated student can diagnose the failure and modify the prompt accordingly when AI fails to provide a useful response that matches his needs. Prompt revisions are necessitated when responses from GenAI do not meet users' requirement. This is also dependent on many factors. A research conducted by Ana et al., 2024 on pre and in-service teachers from October 2024 till January 2025 revealed the correlation among

strategic variables responsible for the adoption of GenAI and prompt techniques in Spain. Evidence from the study objectives through its participants showed that confidence level of trainers forms a pivotal metric for AI adoption in learning among other variables like relevance with instruction, social impact and barriers.

Carrasco-Sáez et al., (2025) developed a methodological framework of four components to aid the prompting techniques among high education students in Chile for qualitative results. As part of their methodology was the exploratory analysis of factors which guided the grouping of adopted techniques with a correlation and regression analysis of potential causal links between techniques and quality of responses. Positive and negative factors correlations were recorded which informed their final conclusion of adopting a guided context strategy for enhancing the quality of response derived from a GenAI model. Learning techniques have been proposed by different studies and the taxonomy for prompt revision is being studied for further regulatory actions (Giray, 2023; Holmes, 2022). The findings of Liu et al., (2026) from their review of existing prompting techniques led to the innovation of a foundational framework for better prompts design and construction.

A multimodal channel data was utilized to enhance self regulated techniques (Azevedo et al., 2025). Related methods adopted in literatures are constraining to word length limits, sample provision, decomposing complex questions to simpler versions and instructing ChatGPT to act like an expert in a particular field. This later technique is commonly adopted by mid-A1 specialist and frequent users. Learners who engage in prompt revision demonstrate stronger metacognitive awareness and critical thinking skills (Zhai, 2022). Each adopted strategy reflects a specific users level of cognition as observed by Rudolph et al. (2023). Most students now prefer to use GenAI for brainstorming rather than copying responses verbatim. The inclusion of machine learning nowadays helps to derive meaningful insight from students' data. A demonstration of this inclusion was performed by Alalawi et al., (2025) when they integrated pedagogical knowledge for performance prediction. The result generated revealed a 18% increase as compared with statistical model. The findings of Kubsch et al., 2022 buttress the effectiveness of learning analytics to aid in tracking learners' progress with supportive adaptive lessons. Synthesis of findings from various studies suggest patterns with evidence that machine learning models increase predictive accuracy and self regulated learning through AI support. However,

these studies did not link prompt revisions to measurable learning gains of the students nor provided a no-code tool like orange for assessing learning gains through educational logs. This study fills this gap by providing an open source Orange workflow that computes an SRS from ChatGPT logs and predicts learning gains.

Methodology and study design

A predictive- correlation and quantitative design was adopted for this study. Both the Pre-test and post-test were conducted on a single group structure. This design approach was chosen because it examined the correlation between revision behaviours and learning gain without variable manipulation. Our single group structure fits this study as it aimed at teaching AI- assisted learning techniques to students through prompt revision.

Participants description and context

A total of 100 undergraduate students from the Federal College of Education, Abeokuta, Nigeria were recruited for the study through purposive sampling. This sampling technique ensured that only students who were engaged in the structures ChatGPT exercise were selected. The participants were mainly of a mixed discipline, combining Computer Science with other science subjects including Mathematics, Physics and Integrated Science. Participants' description is provided in Table 1. They all had prior basic computer literacy skill and had used ChatGPT at least once before. Inform consent for all the participants was gotten for the ChatGPT log analysis after a 2-week elective course on AI-Assisted Learning Strategies.

Table 1. Demography information of Participants

S/n	Department	Level	Number of students	Total
1.	Mathematics	100	12	35
		200	13	
		300	10	
2.	Physics	100	10	30
		200	10	
		300	10	
3.	Integrated Science	100	10	35
		200	15	
		300	10	
Total				100

Course structure description

The course ran for 10 days (weekdays only) involving all the participants. The first day exercise began with a 20-item pre-test multiple choice questions to assess participants' baseline knowledge without ChatGPT usage and an identical post-test of 20-item multiple-choice tests on general study skills and basic AI literacy was conducted after the course.

Students had other exercises requiring their response with access to ChatGPT to aid their responses with log time for each student activity recorded on a GoogleSheet for all the days with their derived responses. Exercises were completed individually using their ChatGPT account. Each exercise was allocated a maximum of 10 minutes.

Table 2: Schedule of Course Description

Day	Activity	Description
1	Pre-test	20 MCQs (baseline knowledge)
2	Exercise 1	"Explain the water cycle to a 10-year-old"
3	Exercise 2	"Debug this Python code: [a sample function]"
4	Exercise 3	"Critique this argument for logical fallacies: [selected short text]"
5	Exercise 4	"Plan a 15-minute activity to teach fractions"
6-8	(no sessions)	–
9	Post-test	Same 20 MCQs (alternate order)
10	Make-up and debrief	–

Instruments and data collection

A structured self-report log form administered via Google Form was used for data collection. The form comprised of sections designated as Student ID, Exercise number, Attempt number, Prompt text, Usefulness rating, Response time and AI response summary.

Post-test and Pre-test were developed and validated by three subject matter experts. Collected scores from the multiple choice tests covered the study scope and basic AI literacy skills.

For each exercise, students used a Google Form to record their anonymized student ID, exercise number (1–4), attempt number (first prompt followed by up to two revisions if usefulness was

rated ≤ 2), the prompt text copied from ChatGPT, a usefulness rating from 1 (not useful) to 5 (very useful), and a self-reported response time in seconds using a stopwatch, while a research assistant manually logged the AI response summary to verify usefulness ratings. This process yielded 400 initial attempts (100 students $\times$ 4 exercises), with low usefulness cases (52% of attempts) generating additional revisions (mean 1.6 per such case), resulting in a final dataset of 2083 rows as shown in Figure 1. Pre-test and post-test scores were initially recorded separately before merging with the student generated dataset, and learning gain was calculated using equation 1.

$$\text{learning gain(lg)} = \frac{(\text{Post Test score} - \text{Pre Test Score})\%}{(100 - \text{Pre Test Score})\%} \quad (1)$$

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
student_id	exercise_number	attempt_num	prompt_text	usefulness_rating	revision_type	persistence	rev_weight	recovery_speed	pre_test	post_test	learning_gain	SRSlearning_gain_percent	NORMrev_weight	NORMpersistence	NORMrecovery_speed	SRS	High_gain
S001	1	1	Water	0.750	0.0	0.0	0	34	67	82	15	0.45455	0.00	0.0	0.03855	0.771	0
S001	2	1	Debug	0.5	0.333	0.0	1	14	78	87	9	0.40909	0.0303	0.0	0.01587	2.1354	1
S001	3	1	Evaluate	0.5	0.0	0.0	0	21	73	96	23	0.85185	0.00	0.0	0.02381	0.4762	0
S001	4	1	Hands-	0.750	0.0	0.0	2	8	73	82	9	0.33333	0.0606	0.0	0.00907	3.8174	1
S002	1	1	Describe	0.5	0.0	0.0	1	52	66	73	7	0.20588	0.0303	0.0	0.05896	2.9972	1
S002	2	1	Fix code	0.5	0.0	0.0	0	21	65	87	22	0.62857	0.00	0.0	0.02381	0.4762	0
S002	3	1	Find	0.0	0.0	0.0	0	14	66	76	10	0.29412	0.00	0.0	0.01587	0.3174	0
S002	4	1	Design	0.750	0.0	0.0	2	5	78	88	10	0.45455	0.0606	0.0	0.00567	3.7494	1
S003	1	1	Water	0.250	0.667	0.5	2	26	75	87	12	0.48	0.0606	0.5	0.02948	14.2256	1
S003	1	2	Water	1.0	1.0	0.5	3	122	75	89	14	0.56	0.0909	0.5	0.13832	18.2204	1
S003	2	1	Debug	0.5	0.0	0.0	0	12	75	90	15	0.6	0.00	0.0	0.01361	0.2722	0
S003	3	1	Analyze	0.750	0.0	0.0	0	10	75	89	14	0.56	0.00	0.0	0.01134	0.2268	0
S003	4	1	Plan	1.0	0.0	0.0	1	15	67	89	22	0.66667	0.0303	0.0	0.01701	2.1582	1
S004	1	1	Describe	0.750	0.0	0.0	0	8	86	88	2	0.14286	0.00	0.0	0.00907	0.1814	0
S004	2	1	Find	0.250	0.0	0.5	2	66	89	98	9	0.81818	0.0606	0.5	0.07483	15.1326	1
S004	2	2	Find	1.0	0.667	0.5	2	66	86	88	2	0.14286	0.0606	0.5	0.07483	15.1326	1
S004	3	1	Evaluate	0.5	0.0	0.0	0	12	86	88	2	0.14286	0.00	0.0	0.01361	0.2722	0
S004	4	1	Activity	0.750	0.0	0.0	3	10	86	88	2	0.14286	0.0909	0.0	0.01134	5.6808	1
S005	1	1	Describe	0.0	0.0	1.0	1	17	65	85	20	0.57143	0.0303	1.0	0.01927	22.2034	1
S005	1	2	Describe	0.250	0.333	1.0	1	17	65	85	20	0.57143	0.0303	1.0	0.01927	22.2034	1
S005	1	3	Describe	1.0	1.0	1.0	0	17	65	85	20	0.57143	0.00	1.0	0.01927	20.3854	1
S005	2	1	Fix code	0.5	0.0	0.0	0	0	65	85	20	0.57143	0.00	0.0	0.000	0	0
S005	3	1	Evaluate	0.750	0.0	0.0	1	10	65	85	20	0.57143	0.0303	0.0	0.01134	2.0448	1
S005	4	1	Hands-	1.0	0.0	0.0	0	12	65	85	20	0.57143	0.00	0.0	0.01361	0.2722	0
S006	1	1	Explain	0.250	0.0	0.5	2	15	65	80	15	0.42857	0.0606	0.5	0.01701	13.9762	1
S006	1	2	Explain	0.750	0.333	0.5	3	35	65	87	22	0.62857	0.0909	0.5	0.03968	16.2476	1
S006	2	1	Find	0.5	0.0	0.0	2	84	65	87	22	0.62857	0.0606	0.0	0.09524	5.5408	1
S006	3	1	Identify	0.750	0.0	0.0	1	10	65	76	11	0.31429	0.0303	0.0	0.01134	2.0448	1
S006	4	1	Activity	0.750	0.0	0.0	0	0	65	67	2	0.05714	0.00	0.0	0.000	0	0
S007	1	1	Water	0.750	0.0	0.0	0	12	86	87	1	0.07143	0.00	0.0	0.01361	0.2722	0
S007	2	1	Fix code	0.750	0.0	0.0	0	14	86	87	1	0.07143	0.00	0.0	0.01587	0.3174	0
S007	3	1	Analyze	0.250	0.0	0.5	1	50	86	87	1	0.07143	0.0303	0.5	0.05669	12.9518	1
S007	3	2	Analyze	1.0	0.667	0.5	2	50	86	87	1	0.07143	0.0606	0.5	0.05669	14.7698	1
S007	4	1	Design	0.5	0.0	0.0	0	12	86	88	2	0.14286	0.00	0.0	0.01361	0.2722	0

Figure 1. Student A.I learning gain dataset collection sample

Coding prompt revision behaviour and self-regulation score (SRS) construction

For the coding of prompt revision behaviour, a developed prompt taxonomy (Liu, 2026) with minor adaptations was adopted where each revision (attempt > 1) was classified by comparing the prompt text to its parent prompt using an Orange Python Script widget as shown in Figure 2. The taxonomy included six categories with their

description and assigned weights tabulated in Table 3. Two researchers independently coded a random 10% sample, achieving substantial agreement (kappa = 0.84), and the remaining rows were coded automatically using keyword matching and length ratio heuristics.

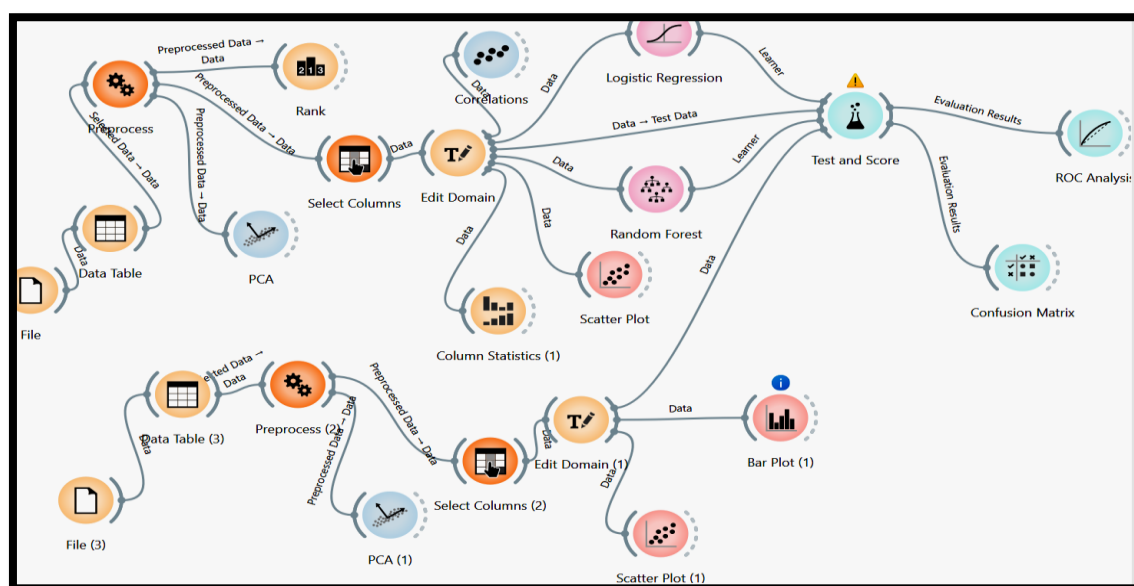


Figure 2. Implementation workflow and architecture of the model

Table 3 Adapted Prompt Revision Taxonomy and Weighted Scores

Revision Type	Definition	Weight
No revision (same prompt)	Identical or nearly identical text (cosine similarity >0.95)	0
Shallow revision	Minor wording change, no new information	1
Constraint addition	Adds length, audience, format, or step-by-step instructions	2
Role/format change	“Act as...”, “Pretend you are...”, “In the style of...”	2
Decomposition	Splits question into sub-questions (e.g., “First... then...”)	3
Example driven	Provides a concrete example of desired output	3

For the self-regulation score (SRS) we formulated and computed three components namely, average revision weight (RevW), Persistence (Pers) and Recovery Speed (Rec). The (RevW) describes the mean weight of all revisions across exercises (where first usefulness ≤ 2), persistence (Pers) shows the average number of attempts per low usefulness exercise, normalized 0–1), and recovery speed (Rec,) details the inverse of time to first substantive revision with weight ≥ 2 , normalized 0–1 and inverted so higher values indicate faster recovery. The final SRS was calculated using equation 2 and all components normalized to [0, 1] using min-max scaling across students, as implemented in the Orange workflow.

$$\text{Final SRS} = 0.6 \times \text{RevW_norm} + 0.2 \times \text{Pers_norm} + 0.2 \times \text{Rec_norm} \quad (2)$$

Implementation workflow description and predictive modelling for learning gain

For The implementation workflow commences with the selection of a File widget that

loads the raw student interaction CSV. This data is further passed to a Python Script widget, which classifies each prompt revision using heuristic rules based on keyword matching and length ratio, outputting a new numeric column called Revision_weight.

Next phase introduces an Aggregate Python widget to group the students' data for the computation of the Self Regulation Score comprising the normalized average revision weight (RevW), persistence (Pers), recovery speed (Rec), and the composite SRS. Finally, a Join widget merges these student level SRS values back with each student's pre-calculated learning gain. The implementation of the predictive learning gain model is also described in Figure 2.

Furthermore, a visualization of the relationship between the SRS and the learning gains is performed using a scatter plot diagram that fits a regression line. The introduction of a correlation widget aids in computing the Pearson's value used in quantifying the association strength. A newly created column is added to sets the binary target

variable (High-gain) based on the learning gain score while the Test and Score widget trains and evaluates a Random Forest classifier (100 trees, max depth 5) using 10-fold cross-validation stratified by student and compared with a Logistic Regression as a baseline comparison learner. Finally, the AUC and ROC curve are generated, and the Confusion Matrix reports precision, accuracy and recall scores, providing a complete evaluation of how well students' revision behaviours predict high learning gains. The high learning gain score is defined by scores in the top quartile precisely with values ($SRS \geq 0.41$).

Results and Observation Analysis

The analyses of the total 1,620 interactions uncovered numerous prompt revision patterns. It showed that 768 (74.01%) records derived were first attempts with rated useful (≥ 3). The remaining 421 (25.99%) records were low usefulness (≤ 2),

generating 586 revisions. Among the revisions recorded, 24% were shallow, 41% constraint addition, 18% role/format, 12% decomposition, 7% example driven. Mean persistence (attempts per low usefulness exercise) was ($M = 1.46$, $SD = 0.47$).

Mean recovery speed (time to first substantive revision) was 22 seconds ($SD = 1.46$). Pre-test scores ranged from 30% to 95% ($M = 67.66\%$, $SD = 0.17$). Post-test scores ranged from 44% to 98% ($M = 76.62\%$, $SD = 0.12$). Mean learning gain was 8.96 ($SD = 0.73$). The top quartile (high gain) threshold was 0.41. A critical look at the interrelation between the Self-Regulation Score (SRS) and learning gain shows some notable correlations as recorded in Table 4. It provided answer to the first research question (To what extent does Self-Regulation Score (SRS) derived from prompt revision behaviour correlate with learning gain).

Table 4: Pearson Correlations between SRS Components and Learning Gain

S/n	Component	Correlation with learning gain(r)	p-value	Strength	Interpretation
1.	RevW	0.44	<0.001	Moderate	Deeper revision helps but not dominant
2.	Pers	0.56	0.002	Moderately strong	Persistence matters most individually
3.	Rec	0.28	0.005	Weak	Speed is least important

Persistence emerged as the highest individual predictor (0.56) followed by revision weight ($r = 0.44$) and recovery speed ($r = 0.28$) respectively. This suggest that students who tried repeatedly after receiving a poor ChatGPT response ultimately learn more despite the quality of their revision strategies. This resonates with the Zimmerman's (2019) theory indicating that persistence is a core behavioural component of self-regulated learning.

The moderate correlation for revision weight indicates that *strategic* revision achieved by

adding constraints, changing roles, or decomposing questions is beneficial, but its effect is increased when combined with persistence. The weak correlation for recovery speed suggests that rushing to revise quickly does not necessarily improve learning rather a careful and thoughtful revision (even if slower) may be more valuable. Most importantly, the composite SRS ($r = 0.87$) depicted in Figure 3 reflects the positive trend and demonstrates that combining these three behavioural indicators produces a far stronger prediction than any single component alone.

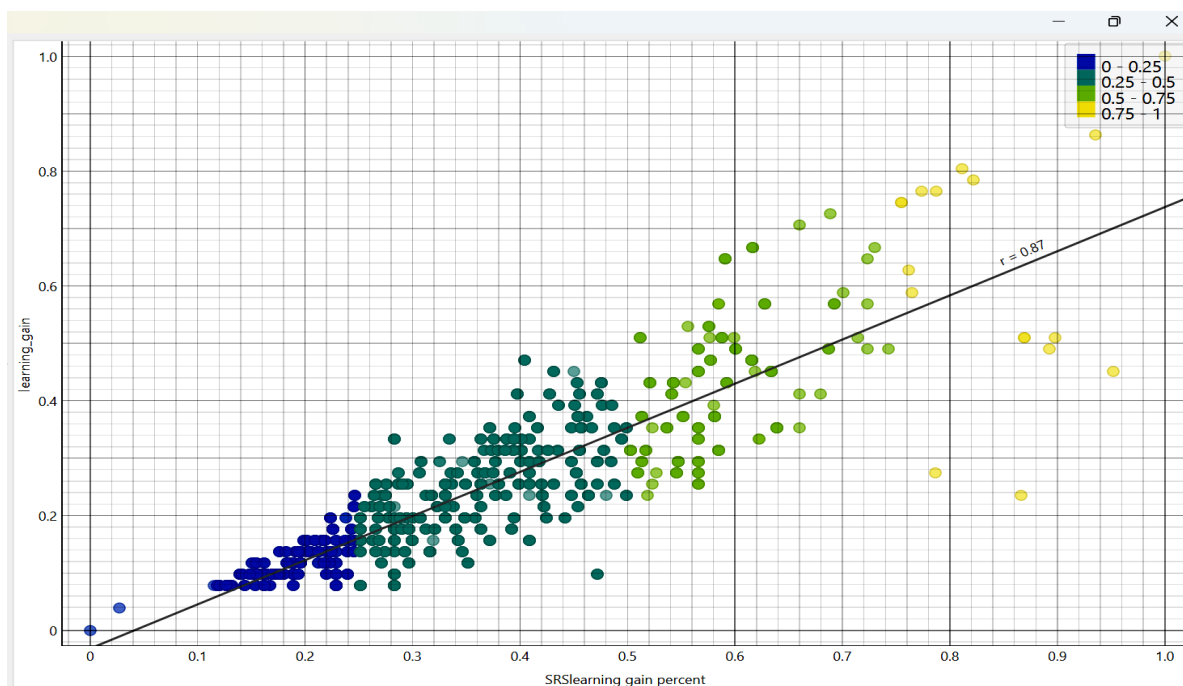


Figure 3. Correlation map of SRS with learning gain showing a positive trend.

To evaluate the performance of revision features on learning gain, Random Forest was trained and tested on subset of the entire dataset (488 instances). Its outputs were compared with the Logistic Regression classifier. The motive was to provide response to the second research question (Can a model using only revision features (no prior grades) predict which students will achieve high learning gain?). Results from the model shown in Figure 4 indicate that Random Forest is a better

classifier than the baseline Logistic Regression for all the evaluated metrics. An AUC improvement score of 25.8% was achieved by Random Forest over its baseline model with 59.6%, indicating that Logistic Regression ranks slightly better than performing a random guess (50%) for classifying selected high-gain and low-gain students. The 75.6% accuracy score pulled by Random Forest from stratifying students by their learning gain also exceeds Logistic Regression by 16.2%.

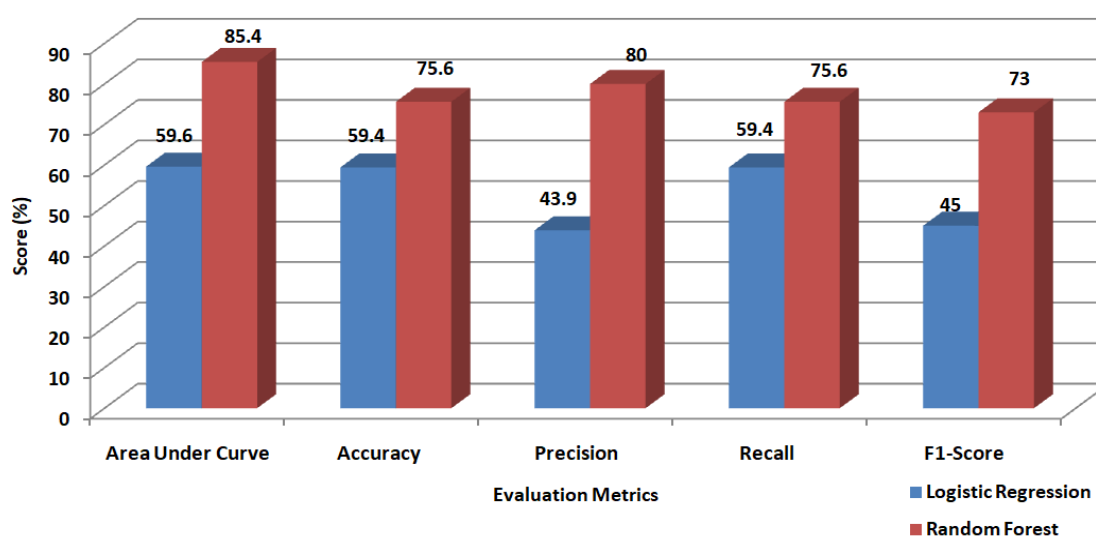


Figure 4: Comparative analysis of the Random Forest with the baseline

The number of students, grouped as high gainers who achieved high gain was 80.0% for

Random Forest, compared to Logistic Regression with 43.9%. Random Forest classifier also

successfully identified three-quarter of correctly identified high-gain students as displayed from its recall (sensitivity) score of 75.6%. The recall and precision balancing metric (F1-score) was 73.0% and 45.0% for Random Forest and Logistic Regression respectively. Average revision weight contributed 51% of the model's predictive power, students persistence rate and recovery speed also contributed 29% and 20% respectively based on their feature

analysis. This confirms the earlier correlation findings of what a student revises (depth) matters most, followed by how many times they persist, and while how quickly they revise provides complementary but secondary value. Confusion Matrices and Receiver operating cost (ROC) graph obtained on the randomly selected test dataset comprising 488 instances (approximately 30% of the collected dataset) are displayed in Figures 5 and 6.

		Predicted		Σ
		0	1	
Actual	0	83	112	195
	1	7	286	293
Σ		90	398	488

Random Forest

		Predicted		Σ
		0	1	
Actual	0	1	194	195
	1	4	289	293
Σ		5	483	488

Logistic Regression

Figure 5. Proportion of Predicted Random Forest and logistic regression.

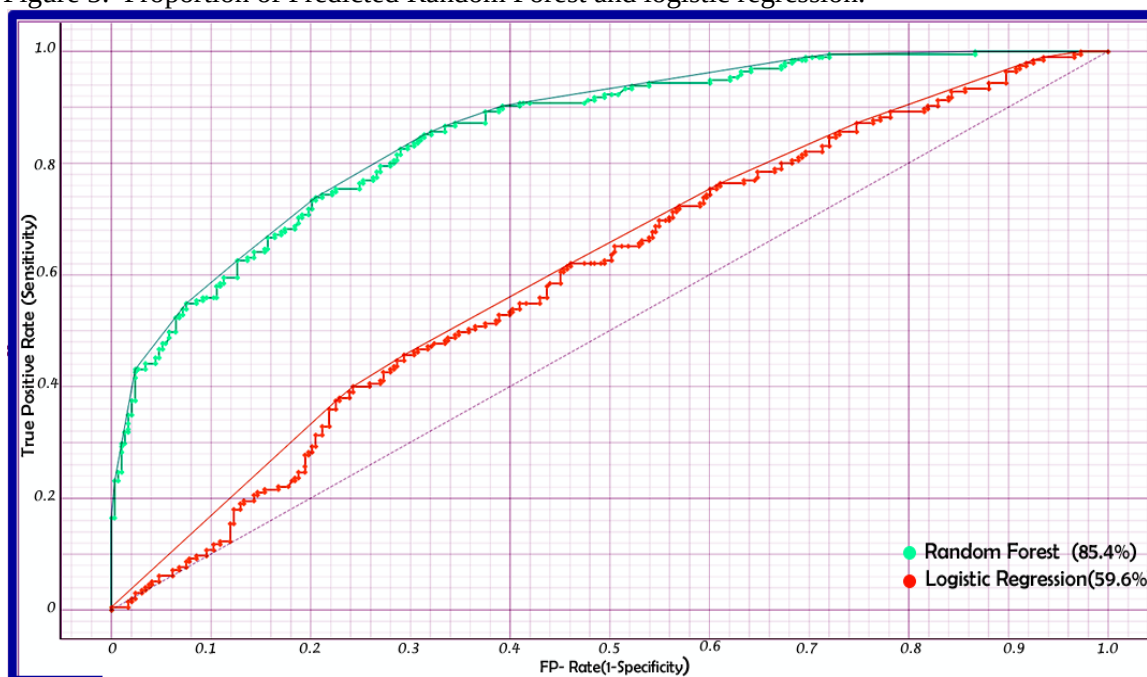


Figure 6. Receiver operating Curve for the developed model.

From the result in Figure 5, out of 293 actual high-gain students, the Random Forest correctly identified 286 (True Positives) and missed only 7 (False Negatives). This defines the model target as it successfully captures 98 out of every 100 students who will achieve high learning gains and only 83 were correctly classified as low-gain, while 112 were incorrectly flagged as high-gain (False Positives). The ROC curve in Figure 6 also graphically confirms the superiority of the Random

Forest model. The Random Forest curve rises steeply towards the top-left corner, indicating high true positive rates (sensitivity) across a range of false positive rates. The area under this curve (AUC) is 85.4%, meaning the model has an 85.4% probability of correctly ranking a randomly selected high-gain student above a randomly selected low-gain student. The AUC score of 59.6% achieved by the Logistic Regression reveals that its performance is more of a random guess than actual prediction. Evidence from

the gap between the two curves, indicates that capturing non-linear relationship from data are significant for predicting learning gains. Table 5

presents the summaries of key findings based on the research questions and feature components ranking.

Table 5: Findings aligned with the Research questions

S/n	Research Question	Key findings	Statistical Evidence
1.	Does SRS correlate with learning gain?	Strong positive correlation	$r = 0.87$ $p < 0.001$ $r^2 = 0.76$
2.	Can revision features predict high learning gain?	Random forest outperforms baseline	AUC = 85.4% Recall = 75.6% Precision = 80.0% Accuracy = 75.6% F1-score = 73%
3.	Feature Importance	Revision weight dominates	RevW = 51% Pers = 29% Rec = 20%

Discussion

Both research questions have been provided with responses (generated results) from this study. Findings reported in Table 5 clarify the inclusion of prompt revision techniques for students learning gains. A combination of revision depth, persistence, and recovery speed which formed the composite Self-Regulation Score (SRS) correlated strongly with the objective learning gain ($r = 0.87$, $p < 0.001$). This further suggests that for better prediction, it is advisable to combine multiple behavioural factors rather than relying on a singular component. The precision and AUC scores of 80% and 85.4% respectively obtained in this study, reflects the Random Forest model's correct ranking and discriminative ability for both high and low-gain students. This evidence supports its adoption for supporting teaching needs and developing accurate learning diagnostic system.

A comparison of the study's correlation score with similar study correlation range ($r \approx 0.4$ – 0.6 ; Zimmerman, 2019) showed that log-based behavioural indicators capture self-regulation more directly than subjective questionnaires. The addition of constraints to prompts has a positive effect on the responses from this study. It therefore aligns with Liu et al., (2025) findings that constraint addition enhances task completion. This empirical finding recommends that students be taught how to add constraints to prompts.

Actionable implications for classroom practice from this study include equipping educators with skills to accurately flag low SRS students in a course. This may necessitate individualized training on prompt

revision techniques for improvement. The problems of privacy concerns highlighted in literature are also eliminated. Student's data need not leave the institution infrastructure, their data are utilized locally within an Orange installed personal computer. This approach aligns with emerging ethical frameworks for student data utilization in AI-enhanced learning environments.

Limitations and future work

Limitations from these points of views are considered. Self-regulation may develop or decay over time. Therefore, the duration of two weeks utilized in this study limits its generalisability to longer academic semesters. A full semester in higher institutions of learning may last for months as approved by the institution senate.

The self-reported usefulness ratings from students, without a verified prior AI experience provide imbalance assessment of their skills. In a bid to avoid been tagged as incompetent, students may provide a false report of their prompting skills. Samples for the study were undergraduate students. The results may not replicate for students in secondary schools or technical institutions. The single revision taxonomy adopted may misclassify complex revisions. Larger samples of the test dataset may improve generalisability than the utilized test instances ($n = 488$).

Addressing these limitations in future work will involve engaging in a longitudinal study for a full semester. This could aid in determining if SRS improves with prompt engineering techniques. It would also check if the improvement can be

sustained for an extended duration. Embedding the Orange workflow into a real-time system's dashboard will be beneficial to both instructors and institution management. Students with SRS below a given threshold can be easily identified for remedial actions. Replication of the study across secondary schools, universities, vocational and technical colleges should also be carried out for generalization.

Conclusion

An evidence-based and privacy preserving method for using prompt revision behaviour to predict student self-regulation is provided in this study. Self-Regulation Score (SRS) correlates strongly with learning gain ($r = .87$). The utilized machine learning model performance is also recommendable for similar research. The Orange workflow reduces the educators' task in implementing rigorous data driven decision processes. It offers a no coding framework to support students learning. As generative AI is becoming useful in digital classrooms, it is required that students and instructors are well equipped with prompting skills to achieve their desired learning outcome. This study offers a validated step in that direction.

References

- Alalawi, K., Athauda, R., & Chiong, R. (2025). An extended learning analytics framework integrating machine learning and pedagogical approaches for student performance prediction and intervention. *International Journal of Artificial Intelligence in Education*, 35, 1239–1287. <https://doi.org/10.1007/s40593-024-00429-7>
- Azevedo, R., & Gašević, D. (2025). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, 154, 108563. <https://doi.org/10.1016/j.chb.2024.108563>
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.61969/jai.1337500>
- Carrasco-Sáez, J. L., Contreras-Saavedra, C., San-Martín-Quiroga, S., Contreras-Saavedra, C. E., & Viveros-Muñoz, R. (2025). Analyzing Higher Education Students' Prompting Techniques and Their Impact on ChatGPT's Performance: An Exploratory Study in Spanish. *Applied Sciences*, 15(14), 7651. <https://doi.org/10.3390/app15147651>
- Chang, E. Y. (2023). Prompting large language models with the Socratic method. *arXiv*. <https://doi.org/10.48550/arXiv.2303.08769>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elugbadebo O.J and Johnson (2020). Computer usage proficiency towards pedagogical knowledge and learning improvement. *Ukrainian Journal of Educational Studies and Information Technology*, Bogdan Khmelnytsky Melitopol State Pedagogical University, Ukraine., Vol. 8, Issue 4, pp.52-66, DOI:10.32919/uesit.2020.04.05
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT in higher education. *Innovations in Education and Teaching International*, 61(3), 1–12. <https://doi.org/10.1080/14703297.2023.2195846>
- Giray, L. Prompt engineering with ChatGPT: A guide for academic writers. *Ann. Biomed. Eng.* 2023, 51, 2629–2633.
- Holmes, W., Miao, F., Ahn, J., & Brunskill, E. (2022). *AI and education: Guidance for policy-makers*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- Johnson, F. T., & Vincent, O. R. (2020). An intelligent system for student placement using fuzzy logic. In *Proceedings of the Fostering IT Systems in Effective Realization of 4th Industrial Revolution*. 100-109. Retrieved from <https://www.researchgate.net/anintelligentsystemforstudentplacement>
- Kubsch, M., Czinczel, B., Lossjew, J., Wyrwich, T., Bednorz, D., Bernholt, S., Fiedler, D., Strauß, S., Cress, U., Drachsler, H., Neumann, K., & Rummel, N. (2022). Toward learning

- progression analytics: Developing learning environments for the automated analysis of learning using evidence-centered design. *Frontiers in Education*, 7, 981910. <https://doi.org/10.3389/feduc.2022.981910>
- Liu, Y.-Y., Zheng, Z., Zhang, F., Feng, J.-C., Fu, Y.-Y., Zhai, J.-D., He, B.-S., Zhang, X., & Du, X.-Y. (2026). A comprehensive taxonomy of prompt engineering techniques for large language models. *Frontiers of Computer Science*, 20(3), 2003601. <https://doi.org/10.1007/s11704-025-50058-z>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence*, 3, 100070. <https://doi.org/10.1016/j.caeai.2022.100070>
- Panadero, E. (2022). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 13, 861511. <https://doi.org/10.3389/fpsyg.2022.861511>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Viberg, O., Khalil, M., & Baars, M. (2023). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. *Educational Technology Research and Development*, 71(3), 987–1013. <https://doi.org/10.1007/s11423-023-10184-0>
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4312418>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2