SelfGauge: An Intelligent Tool to Support Student Self-assessment in GenAI-enhanced Project-based Learning

Chengbo Zheng
cb.zheng@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Zeyu Huang
zhuangbi@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Shuai Ma
shuai.ma@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Xiaojuan Ma
mxj@cse.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

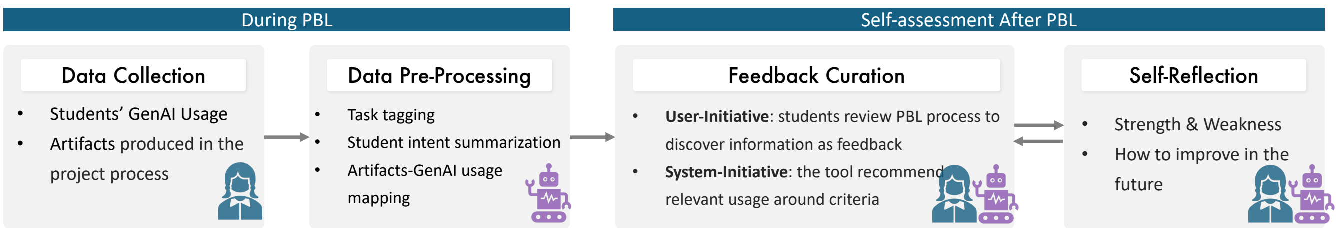

Figure 1: SelfGauge supports students self-assessment through four steps.

ABSTRACT

Project-based learning (PBL) involves students tackling real-world problems and creating artifacts. With the rise of generative AI (GenAI) tools, assessing students in GenAI-enhanced PBL is challenging. To address this, we designed SelfGauge, a tool that supports student self-assessment by analyzing their GenAI usage and project artifacts. It helps students define criteria, seek feedback, and reflect on their performance, promoting continuous self-improvement.

CCS CONCEPTS

• **Applied computing** → **Education**; • **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

AI in Education, Project-based Learning, Reflection

ACM Reference Format:
Chengbo Zheng, Zeyu Huang, Shuai Ma, and Xiaojuan Ma. 2024. SelfGauge: An Intelligent Tool to Support Student Self-assessment in GenAI-enhanced Project-based Learning. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3672539.3686338

UIST Adjunct '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0718-6/24/10
https://doi.org/10.1145/3672539.3686338

1 INTRODUCTION

Project-based learning (PBL) is an educational approach where students learn and apply knowledge through projects that address real-world problems and produce artifacts as solutions [2]. Typically, students are situated in authentic environments and can leverage various kinds of technology in PBL [6]. Meanwhile, there has been a surge of generative AI (GenAI) tools on the market, available to the general public, such as ChatGPT [1], Midjourney [2], and Suno [3]. There is evidence that students are increasingly using GenAI in various learning activities, including PBL [3, 9]. This rise in GenAI usage among students raises a new critical question: how to assess students when they are engaged in GenAI-enhanced PBL [9].

Originally, the artifacts produced in PBL served as key signals for assessing students abilities such as critical thinking and creativity [7]. However, since GenAI can contribute to the creation of these artifacts, the reliability of artifacts as assessment material is threatened. To tackle the assessment challenge in GenAI-enhanced PBL, Zheng et al. [9] suggest that collecting and interpreting students' GenAI usage could be a promising direction to explore. For example, students' behavior in not simply accepting GenAI's answers but asking follow-up questions might show their critical thinking skills [9]. In this work, we build on this idea and design a tool to support students' self-assessment based on their GenAI usage together with design artifacts in their projects. We study self-assessment as it is a critical component of PBL due to PBL's self-directed learning nature [2].

[1]https://chatgpt.com
[2]https://www.midjourney.com
[3]https://suno.com

Education researchers have formalized self-assessment as a cyclical process [8]. This process involves three key actions: 1) Determining Criteria: students define what constitutes "good" performance in their learning activities; 2) Feedback Seeking: students seek both external feedback (e.g., from instructors and peers) and internal feedback (e.g., their emotions) to support their assessment; and 3) Self-reflection: students identify their strengths and weaknesses and reflect on how they can improve in the future. Inspired by this model, we designed an intelligent tool, named SelfGauge, to support student self-assessment in GenAI-enhanced PBL. The workflow of SelfGauge is illustrated in Figure 1. Initially, students upload their GenAI usage and artifacts produced during the learning process to SelfGauge, which pre-processes the data for later interaction. Then, following design guidelines on mixed-initiative interaction [4] and reification [1], SelfGauge reifies criteria as interactive objects, supporting students in creating, characterizing, and seeking feedback around these criteria. Through this criteria-centered interaction process, students incrementally build their self-reflection, identifying their strengths, weaknesses, and areas for future improvement.

2 SYSTEM DESIGN

Given that most GenAI tools are based on text-to-X (X=text, image, etc.) models, we formalize students' GenAI usage as follows: $U = [(t_1, o_1), (t_2, o_2), \ldots, (t_n, o_n)]$, where U denotes the GenAI usage, t_i represents text input, and o_i represents the corresponding output. Below, we outline the four steps of SelfGauge (Figure 1).

During PBL, students document their GenAI usage (e.g., through sharing links) and artifacts, in either text or image, within SelfGauge. This data is pre-processed to enable feedback formulation:

- *Task tagging*. Certain tasks are more reflective for students [9]. For instance, analyzing brainstorming and design processes is more insightful than coding in a design project. We use GPT-4o [4] to tag tasks involved in each student-GenAI interaction turn [5] to support this need.
- *Student intent summarization*. Capturing how students seek GenAI help and their treatment of GenAI suggestions is crucial [9]. SelfGauge uses a recursive approach to summarize student intents during GenAI interactions. For each student message t_i, GPT-4o describes the student's intent I_i by considering the context given $[I_1, S_1, I_2, S_2, \ldots, I_{i-1}, S_{i-1}]$, where S_i is a summary of o_i, and the student's attitude, and their GenAI request given t_i.
- *Artifacts-GenAI usage mapping*. To understand GenAI's role in artifact creation [9], SelfGauge maps each student-GenAI interaction t_i, o_i to relevant artifacts. GPT-4o generates text descriptions for artifacts and extracts key entities from t_i, o_i. The minimum embedding cosine distance between these descriptions and entities ranks their relevance. Timestamps of documentation are also considered, with relevance determined by the PBL time period.

After PBL, SelfGauge supports students to review their PBL process and perform self-assessment. Students first create criteria for assessing their performance, such as critical thinking and creativity. SelfGauge then prompts GPT-4o to provide expected student-GenAI interactions ("expectation" in Figure 2) for each criterion. For instance, for "critical thinking," an expectation can be "students examine multiple GenAI suggestions and choose one with reasons." Criterion-expectation examples from Zheng et al. [9]'s work are provided. Students can edit expectations and add new expectations based on their exploration of their PBL data.

[4]https://openai.com/index/hello-gpt-4o/

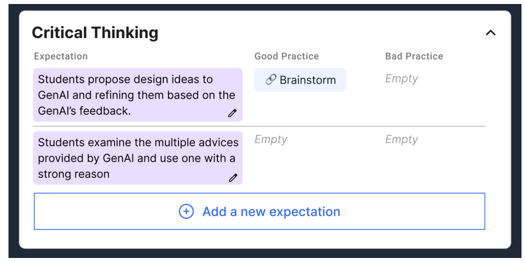

Figure 2: A self-assessment table for "critical thinking".

SelfGauge automatically identifies GenAI usages relevant to student-GenAI interaction expectations using text similarity from the *student intent summary*. For robust retrieval, GPT-4o generates multiple expectation variations to estimate relevance. Relevant usages are recommended for students to review. If students find the usage relevant to the expectations, they can drag and drop it into a self-assessment table (e.g., Figure 2) under the corresponding criterion. The table has three columns: expectation, good practice, and bad practice. Students can place the usage under "good practice" if it meets the expectation, or "bad practice" if not. This interaction prompts students to reflect [8]: understanding their strengths, weaknesses, and potential improvements (expectation).

While the former interaction is mainly system-initiative, SelfGauge also supports user-initiated interactions. Students can browse all GenAI usages indexed by *task tags* and examine uploaded artifacts. Selecting an artifact links it to relevant GenAI usage based on *artifacts-GenAI usage mapping*, allowing students to review GenAI's role in artifact creation. For any usage, they can drag and drop it into a self-assessment table under an existing criterion, adding a new row with an empty "expectation" for the student to fill in.

3 CONCLUSION

Based on a previous co-design study [9] and self-assessment theory [8], we designed SelfGauge to support student self-assessment in GenAI-enhanced PBL. Our next step is to conduct user studies to evaluate SelfGauge. We plan to recruit students in a workshop, where we will introduce design thinking and GenAI tools for designing a mobile application. Afterward, students will work on a week-long design project. They are encouraged to use GenAI tools, which simulate a potential future PBL learning environment [9]. At the end of the week, participants will use SelfGauge to self-assess their work. We will analyze the study data to understand the benefits, scalability, and impacts on students learning outcomes of SelfGauge.

ACKNOWLEDGMENTS

This research was supported in part by the EdUHK-HKUST Joint Centre for Artificial Intelligence (JC_AI) research scheme: Grant No. FB454.

REFERENCES

[1] Michel Beaudouin-Lafon and Wendy E Mackay. 2000. Reification, polymorphism and reuse: three principles for designing visual interfaces. In *Proceedings of the working conference on Advanced visual interfaces*. 102–109.

[2] Phyllis C Blumenfeld, Elliot Soloway, Ronald W Marx, Joseph S Krajcik, Mark Guzdial, and Annemarie Palincsar. 1991. Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational psychologist* 26, 3-4 (1991), 369–398.

[3] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D. Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, Lennart E. Nacke, and Pan Hui. 2024. ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024), 100027. https://doi.org/10.1016/j.chbah.2023.100027

[4] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.

[5] Keming Lu, Hongyi Yuan, Zheng Yuan, Runji Lin, Junyang Lin, Chuanqi Tan, Chang Zhou, and Jingren Zhou. 2023. # instag: Instruction tagging for analyzing supervised fine-tuning of large language models. In *The Twelfth International Conference on Learning Representations*.

[6] Lara Piccolo, Daniel Buzzo, Martin Knobel, Prasanna Gunasekera, and Tina Papathoma. 2023. Interaction Design as Project-Based Learning: Perspectives for Unsolved Challenges. In *Proceedings of the 5th Annual Symposium on HCI Education*. 59–67.

[7] Sarah Sterman, Molly Jane Nicholas, Janaki Vivrekar, Jessica R Mindel, and Eric Paulos. 2023. Kaleidoscope: A reflective documentation tool for a user interface design course. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–19.

[8] Zi Yan and Gavin TL Brown. 2017. A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education* 42, 8 (2017), 1247–1262.

[9] Chengbo Zheng, Kangyu Yuan, Bingcan Guo, Reza Hadi Mogavi, Zhenhui Peng, Shuai Ma, and Xiaojuan Ma. 2024. Charting the Future of AI in Project-Based Learning: A Co-Design Exploration with Students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 94, 19 pages. https://doi.org/10.1145/3613904.3642807