

# How Perceived Learning Difficulties Shift in a Short-Term GenAI-Integrated Design Course

Henry MA <sup>1</sup> & Jinmei HUANG <sup>2</sup>

<sup>1</sup> School of Art & Design, The Hong Kong Polytechnic University, Hong Kong, China

<sup>2</sup> School of Art & Design, Guangzhou College of Commerce, Guangzhou, China

## Abstract

This study examines shifts in students' perceptions of support and learning difficulties within a short-term Generative AI (GenAI)-integrated design course. A pre-post survey design was implemented in a four-week (32 contact-hour) undergraduate design course where GenAI was embedded as a pedagogical component. Matched responses (paired  $n = 36$ ; post  $n = 39$ ) assessed perceived efficiency support, creative support, self-efficacy, and future intention, alongside post-course mechanisms (guidance improvement, workflow formation, iteration intensity) and persistent bottlenecks. Pre-post changes were modest: efficiency support remained high but unchanged, creative support showed a small non-significant increase, and self-efficacy improved with marginal significance ( $p = 0.058$ ). Most students reported an improved ability to guide GenAI and an emerging workflow, while common bottlenecks centred on controllability, consistency, and integration. The findings rely on perception-based measures and should be interpreted as indicators of students' experienced support and difficulty during the course. The results suggest that GenAI compresses early production barriers but shifts difficulty toward higher-order control and synthesis.

**Keywords:** Generative AI, Design education, Workflow formation, Self-efficacy, Learning difficulties

## 1. Introduction

Generative AI (GenAI) is rapidly entering higher education as both a productivity tool and a pedagogical force, prompting institutions to reconsider how learning, assessment, and integrity are designed (UNESCO, 2023; Kasneci et al., 2023; Cotton et al., 2024). In studio-based design education, this shift is especially consequential because learning develops through iterative making, critique, and reflective experimentation rather than one-shot correctness (Schön, 1983). Emerging evidence further suggests that GenAI can support ideation while also introducing new constraints such as fixation, making learning quality harder to infer from outputs alone (Wadinambiarachchi et al., 2024).

This study focuses on how GenAI redistributes difficulty across short-term design learning. Against this backdrop, it remains unclear which student perceptions shift over a four-week instructional cycle and which difficulties persist even when GenAI is integrated into teaching. This study addresses that gap by examining a short undergraduate design course using matched pre-post surveys to track changes in perceived support, strategy formation, and persistent bottlenecks around controllability, consistency, and integration.

To guide this investigation, the study was framed by the following research questions:

- **RQ1:** How do students' perceptions of efficiency support, creative support, self-efficacy, and future intention change from pre to post in a GenAI-integrated design course?
- **RQ2:** What process mechanisms emerge after the course (guidance improvement, workflow formation, and iteration intensity), and how are these mechanisms associated with perceived gains (efficiency gain, quality gain, self-efficacy, and success in realising ideas)?
- **RQ3:** What persistent bottlenecks do students report when using GenAI, and are these difficulties linked to pre-course controllability concerns and post-course self-efficacy differences?

## 2. Methodology

A pre-post survey design was employed to examine short-term shifts in students' perceptions of support and difficulty in a four-week required first-year undergraduate course, *Artificial Intelligence and Visual Communication*, totaling 32 contact hours. The course guided students through most stages of the design-thinking process, from identifying a problem and defining a concept to generating proposals, developing visual outcomes, and producing a final short video. Deliverables included logo design, IP character design, static and animated stickers, static and motion posters, a video script, and an approximately 40-second video. GenAI was integrated throughout these stages as part of ideation, visual exploration, refinement, and production support rather than being introduced as an optional add-on.

The pre-survey was administered after the first class, and the post-survey after the final class. The pre-survey captured students' baseline GenAI experience, expectations, and concerns. The post-survey measured perceived efficiency and task progression, perceived improvement in creative output quality, self-efficacy/controllability in using GenAI, and strategy formation in using GenAI. Perceived alignment is used here as a proximal indicator of students' felt ability to meet course demands. These measures are indirect proxies; therefore,

they should not be read as direct evidence that course learning outcomes were fully achieved.

The analysis tracks changes in perceived support, strategy formation, and difficulty during the course and helps locate where the pedagogy appeared to support progress or where problems remained from students' perspectives. It also collected process indicators such as iteration frequency, typical actions after initial AI outputs, encountered bottlenecks, and coping responses. Responses were matched across time points using the last four digits of student IDs to enable paired comparisons. Paired t-tests were used for comparable pre-post Likert items, while descriptive statistics summarised post-course distributions of strategy formation and bottleneck patterns. Non-significant changes are interpreted descriptively, not as confirmatory evidence, and are considered pedagogically relevant only when they align with post-course process indicators and recurring bottleneck patterns. Multi-select difficulty items were coded into binary variables and visualised to identify persistent challenges. Artifact-based or performance-based assessment was not included in this short paper and remains a priority for future work.

### 3. Results and Discussion

Matched pre-post responses were obtained for 36 students (paired  $n = 36$ ), and 39 students completed the post-survey (post  $n = 39$ ). Overall, the findings indicate modest shifts in students' perceptions of support alongside concentrated bottlenecks in controllability, consistency, and integration. Key quantitative findings across RQ1–RQ3 are summarised in Table 1. Taken together, these results support a cautious interpretation of how GenAI functions in a short-cycle design pedagogy.

Table 1 summarises the full quantitative pattern across the three research questions, including pre-post changes, post-course mechanisms, group differences linked to reported bottlenecks, and descriptive post-course distributions. Figure 1 corresponds to the pre-post comparison in RQ1; Figures 2a and 2b correspond to the post-course mechanism patterns in RQ2; and Figure 3 corresponds to the persistent bottlenecks highlighted in RQ3.

**Table 1.** Key findings on pre-post changes, post-course mechanisms, and persistent bottlenecks (paired n = 36; post n = 39)

| Indicator / comparison                                                                                 | Key result (condensed)                           | p     | Sig. |
|--------------------------------------------------------------------------------------------------------|--------------------------------------------------|-------|------|
| <b>Panel A. Pre-post changes in key perception-based indicators (paired n=36)</b>                      |                                                  |       |      |
| Efficiency support (Likert)                                                                            | Pre 4.14±0.59; Post 4.14±0.87; t=0.00; dz=0.00   | 1     | n.s. |
| Creative support (Likert)                                                                              | Pre 3.81±0.86; Post 3.94±0.95; t=0.78; dz=0.13   | 0.443 | n.s. |
| Self-efficacy (experience→effective use)                                                               | Pre 3.61±0.77; Post 3.89±0.82; t=1.96; dz=0.33   | 0.058 | †    |
| Future intention to use GenAI                                                                          | Pre 4.11±0.62; Post 4.06±0.92; t=-0.39; dz=-0.06 | 0.701 | n.s. |
| <b>Panel B. Post-course mechanisms: guidance and strategy formation (matched n=36)</b>                 |                                                  |       |      |
| Workflow formation ↔ Self-efficacy                                                                     | q=0.559                                          | <.001 | ***  |
| Workflow formation ↔ Quality gain                                                                      | q=0.491                                          | 0.002 | **   |
| Workflow formation ↔ Efficiency gain                                                                   | q=0.468                                          | 0.004 | **   |
| Guidance improvement ↔ Efficiency gain                                                                 | q=0.514                                          | 0.001 | **   |
| Guidance improvement ↔ Workflow formation                                                              | q=0.634                                          | <.001 | ***  |
| Workflow formation ↔ Success in realizing idea                                                         | q=0.443                                          | 0.007 | **   |
| <b>Panel C. Persistent bottlenecks: group differences by reported difficulties (matched post n=36)</b> |                                                  |       |      |
| Pre controllability concern → Post 'hard to integrate'                                                 | Issue=3.53 (n=15) vs No-issue=2.95 (n=21)        | 0.035 | *    |
| Post 'hard to integrate' ↔ Self-efficacy                                                               | Issue=3.60 (n=15) vs No-issue=4.10 (n=21)        | 0.082 | †    |
| Post 'unstable outputs' ↔ Efficiency gain                                                              | Issue=3.90 (n=21) vs No-issue=4.47 (n=15)        | 0.056 | †    |
| <b>Panel D. Descriptive distributions and prevalence (post n=39)</b>                                   |                                                  |       |      |
| Efficiency gain ≥4/5                                                                                   | 71.8% (28/39)                                    |       |      |
| Quality gain ≥4/5                                                                                      | 66.7% (26/39)                                    |       |      |
| Self-efficacy ≥4/5                                                                                     | 64.1% (25/39)                                    |       |      |
| Guidance improved (any)                                                                                | 97.4% (38/39)                                    |       |      |
| Workflow formed (clear/mostly)                                                                         | 79.5% (31/39)                                    |       |      |
| Iterated ≥4 times                                                                                      | 59.0% (23/39)                                    |       |      |
| Bottleneck: unstable / hard to reproduce                                                               | 61.5% (24/39)                                    |       |      |
| Bottleneck: mismatch with intention                                                                    | 53.8% (21/39)                                    |       |      |
| Bottleneck: aesthetics hard to refine                                                                  | 53.8% (21/39)                                    |       |      |
| Bottleneck: hard to edit/integrate                                                                     | 41.0% (16/39)                                    |       |      |

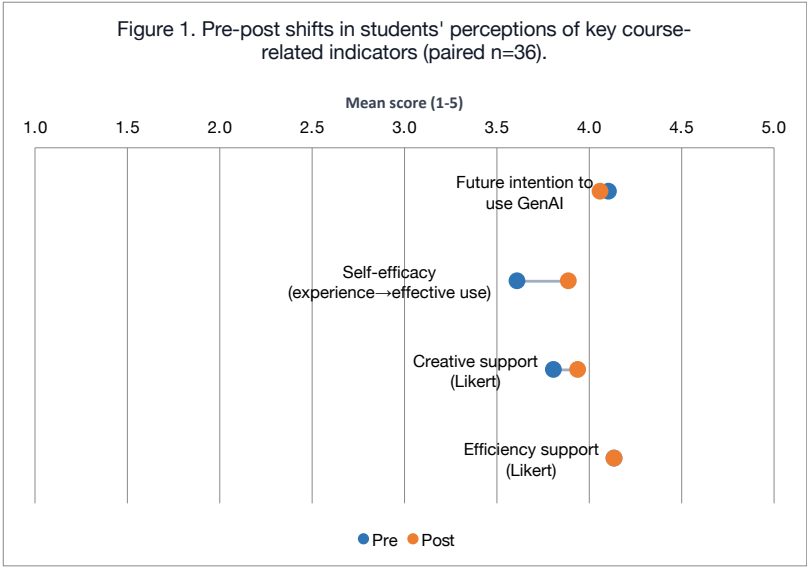
*Note:*

\* Pre-survey n = 43; Post-survey n = 39; matched paired n = 36 (via last four digits of student IDs).

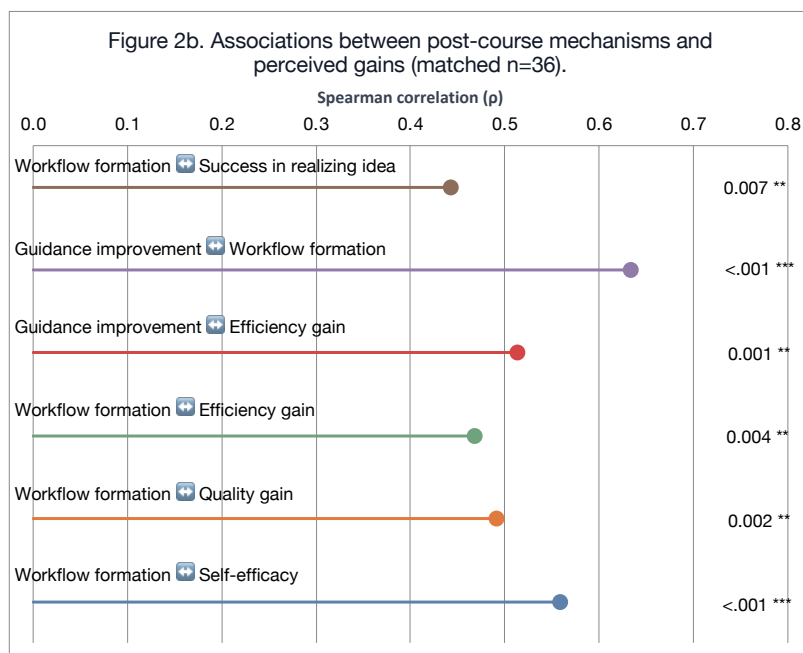
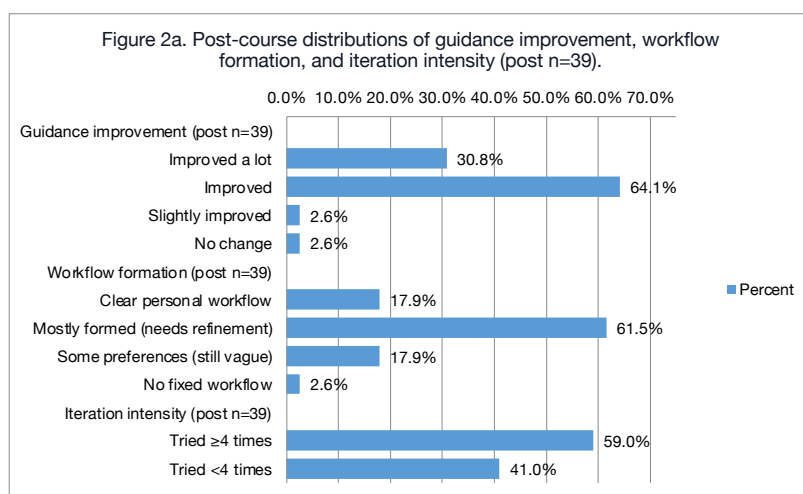
\* Panel A uses paired-samples t-tests (Cohen's dz).

- \* Panel B uses Spearman rank correlations ( $\rho$ ).
- \* Panel C uses Mann–Whitney  $U$  tests (two-tailed).
- \* Panel D reports descriptive percentages only.
- \* Significance codes: \*\*\*  $p < .001$ ; \*\*  $p < .01$ ; \*  $p < .05$ ; †  $p < .10$ ; n.s. not significant.

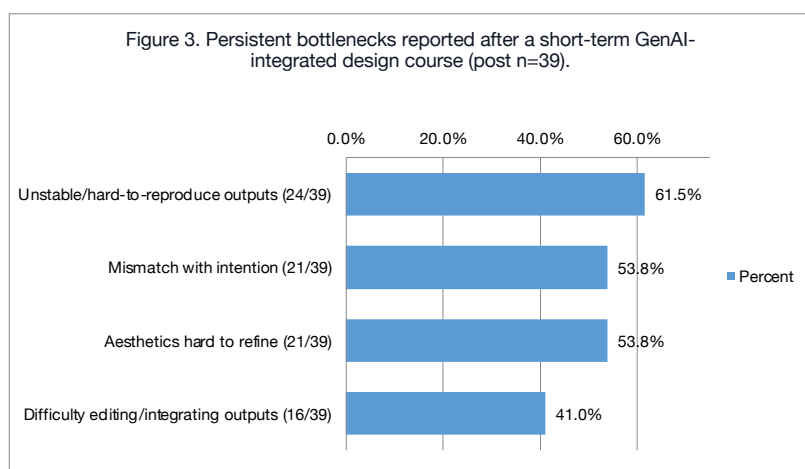
**RQ1 (paired  $n = 36$ ).** Efficiency support remained high but unchanged ( $M = 4.14 \rightarrow 4.14$ ,  $t = 0.00$ ,  $p = 1.000$ , n.s.). Creative support increased slightly ( $M = 3.81 \rightarrow 3.94$ ) but was not significant ( $p = 0.443$ , n.s.). Self-efficacy showed the largest gain ( $M = 3.61 \rightarrow 3.89$ ), reaching marginal significance ( $p = 0.058$ , †). Future intention to use GenAI remained stable ( $M = 4.11 \rightarrow 4.06$ ,  $p = 0.701$ , n.s.). Descriptively (post  $n = 39$ ), 71.8% (28/39) rated efficiency gain  $\geq 4/5$ , 66.7% (26/39) rated quality gain  $\geq 4/5$ , and 64.1% (25/39) rated self-efficacy  $\geq 4/5$ . These pre-post shifts are visualised in Figure 1. The pre-post movement is therefore modest rather than decisive.



**RQ2 (post  $n = 39$ ; associations on matched  $n = 36$ ).** Most students reported an improved ability to guide GenAI (97.4%, 38/39) and an emerging workflow (clear/mostly formed: 79.5%, 31/39). Iteration was common: 59.0% (23/39) reported trying  $\geq 4$  times. Figure 2a displays the descriptive distributions of guidance improvement, workflow formation, and iteration intensity. Mechanism indicators were strongly associated (Table 1, Panel B): workflow formation correlated with self-efficacy ( $\rho = 0.559$ ,  $p < .001$ ), quality gain ( $\rho = 0.491$ ,  $p = 0.002$ ), and efficiency gain ( $\rho = 0.468$ ,  $p = 0.004$ ); guidance improvement correlated with efficiency gain ( $\rho = 0.514$ ,  $p = 0.001$ ) and workflow formation ( $\rho = 0.634$ ,  $p < .001$ ). Figure 2b visualises these associations.



**RQ3 (post n = 39; matched n = 36 for group tests).** Frequent bottlenecks included unstable/hard-to-reproduce outputs (61.5%, 24/39), mismatch with intention (53.8%, 21/39), aesthetics hard to refine (53.8%, 21/39), and difficulty editing/integrating outputs (41.0%, 16/39). Group comparisons showed that students reporting “hard to integrate” had higher pre-course controllability concerns ( $p = 0.035$ , \*) and tended to report lower self-efficacy ( $p = 0.082$ , †) (Table 1, Panel C). Figure 3 visualises the concentration of these persistent bottlenecks.



Across the three research questions, the results point to a clear redistribution of difficulty. Pre-post changes were modest. The clearest signal comes from the combination of workflow formation, improved guidance ability, and concentrated bottlenecks around control, consistency, and integration. This matters because classroom use of GenAI involves more than faster production; students still need to evaluate, direct, and refine what the system generates. The evidence here reflects students’ perceptions of their ability to engage with course demands over a short cycle and should be read as an indirect indication of learning-related experience rather than a direct measure of achieved learning outcomes.

Studio-based design learning traditionally develops through iterative making, critique, and reflective experimentation (Schön, 1983). Difficulty remained present throughout the course, although it shifted from producing options to steering, refining, and integrating outputs into coherent design decisions. This pattern resonates with concerns that GenAI may encourage superficial product completion unless process evidence is made visible and valued (UNESCO, 2023; Xia et al., 2024). The persistent bottlenecks also suggest a more concrete pedagogical agenda. “Unstable or hard-to-reproduce outputs” points to a consistency problem: students may generate useful results once but cannot reliably regenerate or adapt them. “Mismatch with intention” points to weak controllability: students may describe goals broadly but lack methods for specifying constraints, references, and decision criteria. “Difficulty editing/integrating outputs” points to a synthesis problem: even acceptable AI outputs may remain hard to combine, refine, or translate into final design work. These recurring failure points help explain why short-term difficulty remains concentrated in control and integration despite high reported efficiency support.

## 4. Implications

For design educators, the practical implication is to teach control-oriented routines and embed GenAI within a broader design process. Three intervention areas appear especially important. First, **controllability strategies**: prompt iteration with explicit constraints, reference anchoring, and negative instructions can help students move from vague requests to steerable outputs. Second, **consistency and reproducibility routines**: students should document prompt versions, parameter choices, and selection criteria so outputs can be repeated, compared, and refined. Third, **integration and editing workflows**: courses should require students to combine AI outputs with manual editing, layout decisions, material judgment, and reflective rationale so that synthesis remains visible. These strategies respond directly to the main bottlenecks identified here and may better support studio-like learning processes in short-term GenAI use. Future work should triangulate these perception-based findings with artifact-based assessment and rubric-aligned judgments of integration quality and process documentation.

## Acknowledgments

The authors thank the students from a design class at a university in Southern China for their participation in this study.

## Ethical Declaration Statement

The project has been approved by the School of Design of The Hong Kong Polytechnic University (Reference Number: HSEARS20250217010).

## Generative AI Policy

The authors used ChatGPT during the preparation of this work to refine language and check formatting; after utilising the tool, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

## References

- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good?

- On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- UNESCO. (2023). *Guidance for generative AI in education and research*. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., & Velloso, E. (2024). The effects of generative AI on design fixation and divergent thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24, Article 380, pp. 1-18). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642919>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), Article 40. <https://doi.org/10.1186/s41239-024-00468-z>