Reconstructing Skills Assessment in Chinese Vocational Marketing Education under Generative AI

Feng Han*
Zhejiang Business Technology Institute, Ningbo 315012, China
*Corresponding email: hanfeng@zbti.edu.cn

Abstract

Chinese vocational colleges must teach students to work with generative artificial intelligence (AI), but the instruments used to judge skill have not kept pace. In marketing and practice-based programmes, skill is assessed through a finished artefact, and generative tools have weakened the inference from artefact to competence. Drawing on the redesign and judging of a marketing skills contest at one vocational college, this paper identifies three difficulties that now affect such assessment. The artefact no longer reflects the effort behind it, the scoring criteria contain no element for how AI is used, and a team's account of its process cannot be verified. It sets out the grounds on which assessment can be reconstructed, taking constructive alignment, the criteria dimension of authentic assessment, and the levels of AI literacy as its reference points. It proposes four pathways: pairing an open choice of task with hard formal constraints, asking for evidence rather than declarations, bringing oral defence forward from the final round, and writing the evaluation of AI output into course objectives. The argument is that criteria written before these tools arrived cannot settle what good use of them looks like. Part of the mark must be anchored in what a tool cannot produce.

Keywords

Skills assessment, Generative artificial intelligence, Assessment design, Vocational education, Marketing education

Introduction

Chinese policy now asks schools at every level to bring AI into teaching. Reviews of the vocational sector report that the direction is agreed while implementation lags. Much of the discussion so far has concerned how teachers might use these tools to teach and how students might use them to learn. How student work should be judged once the tools are in the students' hands has received less attention, and it is the point at which the other questions become practical [1].

The difficulty is sharpest in practice-based programmes. In marketing, as in design and media, skill has traditionally been inferred from the quality of a finished artefact. A video that is well shot, cleanly cut and clearly subtitled was taken as evidence that its maker could shoot, cut and subtitle. Generative tools weaken that inference. The artefact can now be produced with little of the craft it used to require, so the same evidence no longer licenses the same conclusion about the person who submitted it [2].

Most of the instruments a programme uses are slow to adjust. A programme specification is normally revised on a three-year cycle. Course standards follow it. A skills contest is the exception: Its regulations, deliverables and scoring can be rewritten within a semester. This makes it the fastest instrument a teaching team controls and a reasonable place to work the problem out [3].

This paper asks three questions: Where exactly does skills assessment fail under generative tools, on what grounds can it be reconstructed, and what can a teaching team do within one semester? It draws on the redesign and judging of a marketing skills contest at a vocational college in Ningbo, renamed Digital and Intelligent Marketing Skills in 2026 after a programme of that name entered the national vocational programme catalogue.

Three difficulties under generative AI

The artefact no longer reflects the effort behind it

Smooth camera movement, clean editing, clear subtitles

and a steady voice track were once where effort showed. They are now largely supplied by the tools, and supplied to every team at once. An assessment that continues to reward them is measuring access rather than ability, and it stops separating teams that worked from teams that prompted.

This is not an argument that the artefact has become worthless. It still shows taste, pacing and whether a claim has been thought through. The point is narrower: The technical surface, which used to be the easiest thing to score and the most reliable proxy for work done, has stopped carrying that information [4].

The scoring criteria contain no element for how AI is used

Criteria in use today were written before these tools arrived. The five-element scheme applied in the contest studied here, covering technical quality, practical value, creativity, teamwork and professional conduct, says nothing about generative tools, and neither do the course assessment schemes it was derived from.

Judges nonetheless have views about AI use, and those views enter the mark without passing through any criterion. In the contest studied, written comments returned to the subject repeatedly and in two opposite directions. Teams that showed their process were rewarded for it; entries that looked generated and unexamined were marked down. Because no shared statement of good AI use exists, each judge applies a private one, and the consequence is visible in the scores: The same entry could draw markedly different marks from different judges, enough to move it several places in the ranking. Which judges a team happened to draw mattered more than it should. Nor is this solved by averaging. Judges differed not only in what they valued but in how severely they marked, so an average formed from private and incompatible standards is stable without being meaningful [5].

A team's account of its process cannot be verified

The obvious repair is to ask students to describe how they used AI. The difficulty is that such a description is itself text, and text is what these tools produce most readily. An account of a careful process is no harder to generate than the artefact it describes.

This is not hypothetical. Among the entries judged, several decks described a production process that the team's own video appeared to contradict, claiming live footage that was plainly generated, or on camera appearances that never took place. A declaration, however detailed, is a claim rather than a record, and an assessment that scores declarations will reward the teams that write them best [6].

Grounds for reconstruction

Alignment has to be rebuilt at the design stage

Teaching works when intended outcomes, learning activities and assessment point the same way. This alignment must be built into the design rather than repaired at the marking table. Generative tools disturb the alignment from the assessment end only. The outcome a programme intends and the activity it sets can both remain sound while the evidence linking them quietly changes meaning. It follows that the repair belongs in the design of the task and its criteria. It does not belong in stricter marking of an unchanged task [7].

The criteria dimension is the one that has failed

Authenticity in assessment can be described along five dimensions: the task, the physical context, the social context, the form of the result, and the criteria. A contest of the kind described here performs well on the first four. The task is a real promotional problem, the setting and the social arrangement resemble agency work, and the result is the artefact a client would receive. What has failed is the fifth. Of the five, criteria are also the cheapest to rewrite, since doing so requires no new equipment, no external assessors and no change to the programme specification [8].

Assessment should target evaluation rather than operation

Ng and colleagues divide AI literacy into knowing and understanding, using and applying, evaluating and creating, and ethics [9]. Operating a generative tool sits in the second of these, and it is now widely distributed and quickly learned. Testing what the tool returned, deciding whether it will serve, and taking responsibility for that decision sit in the third, and they are neither widely distributed nor quickly learned. Among the entries judged, competence in operation was common and evidence of evaluation was not. Assessment that still targets operation is aiming at the part of AI literacy that no longer discriminates.

Pathways for reconstruction

Pairing an open choice of task with hard formal constraints

Setting no common case, and requiring each team to choose and justify a real product or service, moves the first demand of the work to task definition. A common case invites a single prompt and produces convergent entries, since teams querying the same model with the same brief receive similar answers.

Open choice alone, however, leaves nothing that can be checked quickly, so it should be paired with constraints that an entry either meets or fails. Three are worth adopting. A fixed length of about thirty seconds, rather than a minimum, turns editing into prioritisation, because generative tools make length cheap and brevity expensive. A landscape format is required so that the work can be shown on screen at the live round, since vertical video suits the platforms where such content is normally watched but presents poorly to a room. A mandatory human voice keeps one channel in which a person must appear in the work, and it corresponds to an entry level occupational task, as short video presenting and live commerce now are.

Asking for evidence rather than declarations

Adding an AI use element to the rubric is the intuitive response to the second difficulty, and it is the wrong one. Such an element can only score what a team says about its own process, and a statement of process is text, which is exactly what these tools produce most readily. The rubric would end up rewarding the teams that write the best account of themselves. What can be asked for instead are artefacts that carry their own traces. A first version submitted beside the final one, with a sentence on what changed and why, is harder to fabricate than a paragraph describing care. So are the prompts that failed, the source of any figure quoted, and a record of the checks a team ran before submitting. These are also more useful to read, because they show the decisions rather than asserting that decisions were made. The shift costs the assessor little. Comparing two versions of one video takes less time than reading a page of prose about how the video was made, and it gives a judge something to point at when explaining a mark [10].

Bringing oral defence forward from the final round

A file can be generated. An answer to an unexpected question, given in front of other people, cannot yet be. Oral defence is therefore the one element of the assessment that generative tools do not reach, and it is commonly placed where it does the least work, namely in a final round after most of the field has been eliminated on files alone. Moving a short oral component into the first stage, even two or three minutes per team with one question that only someone who made the work could answer, converts the strongest available safeguard into the one that filters.

Writing the evaluation of AI output into course objectives

The first three changes are changes to a contest. A contest runs once a year and reaches only the students who enter it, so it can expose the problem without solving it. The change that matters is to the courses.

Only one thing has to change in them. What the student is asked to hand in should move from the output of a tool to a judgement about that output. This does not require a new course, only a rewording of assignments in the courses that already exist.

An assignment that asked students to write a promotional line with AI can instead ask them to generate three, choose one, and say what was wrong with the other two. An assignment that asked for an AI assisted analysis of sales figures can ask for the analysis together with one claim in it that does not hold up, and an explanation of how the student noticed. Where students may quote a market figure produced by a tool, the assignment can require the original source, with no credit if the figure cannot be traced to one.

What these have in common is that the second half of each task cannot be delegated. Anyone who can operate a tool can complete the first half. Only a student who has understood the material can complete the second, and the second half is the part worth marking [11].

Conclusion

The evidence here comes from one contest at one college, described by the teacher who organised it, and the pathways proposed have not been tested against a comparison group.

Within that limit the argument is straightforward. Generative tools have not made skills assessment impossible, but they have broken the link between a finished artefact and the ability of the person who

submitted it. Criteria written before these tools arrived cannot settle what good use of them looks like, and while they stand unrevised, judges will settle it privately and inconsistently. Reconstruction does not require new technology or a new curriculum. It requires that part of the mark be anchored in what a tool cannot produce: a question answered live, a draft that can be compared with its successor, and a judgement the student has to own.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F. (2019) Systematic review of research on artificial intelligence applications in higher education-where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.

[2] Perkins, M. (2023) Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), 1-24.

[3] Li, Y., Cheng, H., Qin, Q. (2025) Evaluations and improvement methods of deep learning ability in blended learning. *International Journal of e-Collaboration (IJeC)*, 21(1), 1-17.

[4] Fyfe, P. (2023) How to cheat on your final paper: assigning AI for student writing. *Ai & Society*, 38(4), 1395-1405.

[5] Sullivan, M., Kelly, A., McLaughlan, P. (2023) ChatGPT in higher education: considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 31-40.

[6] Lancaster, T. (2023) Artificial intelligence, text generation tools and ChatGPT-does digital watermarking offer a solution? *International Journal for Educational Integrity*, 19(1), 10.

[7] Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Gašević, D. (2022) Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075.

[8] Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., Herrera-Seda, C. (2018) Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840-854.

[9] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., Qiao, M. S. (2021) Conceptualizing AI literacy: an exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.

[10] Sotiriadou, P., Logan, D., Daly, A., Guest, R. (2020) The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 45(11), 2132-2148.

[11] Deng, Z., Xiang, H., Tang, W., Cheng, H., Qin, Q. (2024) BP neural network-enhanced system for employment and mental health support for college students. *International Journal of Information and Communication Technology Education (IJICTE)*, 20(1), 1-19.