2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC

Host Sponsor UNC Charlotte

Paper ID #51745

MAKER: Expertise-adaptive GPT-based Manufacturing Adviser using reverse Q&A

Fatemeh Karimi Kenari, University of North Carolina at Charlotte
Mahmoud Dinar, University of North Carolina at Charlotte

Mahmoud Dinar is an assistant professor of mechanical engineering. His main research interest involves integrating AI in multimodal computational frameworks to understand and aid design for manufacturing. His ongoing projects are creating a conversational platform to evaluate and teach manufacturing skills to the future workforce, and solving geometric puzzles to design sustainable and resilient products as a 3D composition of pieces.

Shreyes N Melkote, Georgia Institute of Technology

Shreyes N. Melkote is Professor of Mechanical Engineering at Georgia Tech. His area of teaching and research is advanced manufacturing. He is a Fellow of ASME, SME, and CIRP.

Expertise-adaptive GPT-based Manufacturing Adviser using reverse Q&A

Abstract

The Manufacturing Adviser is a GPT-based question answering testbed for adaptive tutoring in manufacturing. Earlier, we created a question-answer (Q&A) corpus for four manufacturing related topics: process, sub-process, parameter, and design-for-manufacturing. Here, we utilize a Bloom-aligned Q-matrix for evaluating and adapting to learner expertise through reverse Q&A. The Adviser asks questions to gauge user understanding before or during an interaction. By evaluating the specificity, correctness, and reasoning within responses, the system estimates a learner's proficiency and adapts subsequent questions posed at or asked by the users accordingly. This approach supports more effective scaffolding of learning content and enables the Adviser to act as a semi-autonomous tutor rather than a simple question-answering tool. This enhancement aims to make the Manufacturing Adviser more context-aware, and user-friendly to promote the education of and broaden participation in the future manufacturing workforce.

1.Introduction

Manufacturing industry seeks engineers who can operate effectively in evolving production environments. However, manufacturing engineering programs often miss what industry expects from graduates upon entry in workforce-required competencies [1]. Programs are expected to position graduates at the intersection of processes, data, automation, and quality while still retaining the foundational ability to reason about materials, geometry, tooling, and defects [2]. Emerging instructional technologies like Large Language Models (LLMs) can interactively convey concepts that are otherwise difficult to teach through static diagrams or lectures alone. Despite richer simulations and digital artifacts, learners still vary substantially in the depth of explanation and scaffolding they require to interpret instructions and to translate it into correct process and design decisions. AI-enabled tutoring expands what can be visualized and practiced, enabling adaptive, on-demand guidance that helps students at an appropriate cognitive level [3]. LLMs can act as virtual tutoring agents through conversational interaction and support learning through interactive guidance. For instance, Caccavale et al. [4] describe the potential role of LLMs as virtual tutors with iterative clarification and contextualized reasoning for chemical engineering. Yet, there are implementation risks especially in reliable, transparent, and instructionally valid assessment and feedback. Giannakos et al. [5] synthesize opportunities and challenges of generative AI in education, emphasizing the need for safeguards and evaluative rigor, and not mistaking fluent generation for correct or instructionally appropriate support. Intelligent tutoring systems (ITS) have taken heterogeneity in learner capability into account. A review of adaptive ITS for STEM education characterizes adaptive tutoring around learner modeling, personalization, and feedback strategies that respond to student performance which must distinguish between novice-level recall and expert-level decision reasoning [6]. Domain-specific ITS in manufacturing-related training also illustrates that adaptivity can be operationalized around procedural problem-solving and targeted hints rather than generic explanation alone. The CNC Tutor [7] shows the value in diagnosing what the learner is doing wrong and why. A central challenge in adaptive tutoring is deciding what counts as evidence of "understanding". Bloom's revised taxonomy remains one of the widely used educational frameworks for structuring cognitive demand in assessment and instruction. Jackson [8] explicitly connects Bloom's revised taxonomy to the design of prompts and learning activities, emphasizing distinctions among cognitive levels that can be leveraged to make learner proficiency observable through progressively more demanding tasks. Importantly for adaptive dialogue, the paper also discusses "flipping the interaction paradigm" by prompting AI systems to ask learners questions rather than only providing answers. Reverse Q/A is a diagnostic or Socratic probing where the tutor controls question sequencing to surface competence.

To connect question design to principled proficiency inference, cognitive diagnosis research introduces the Q-matrix as a structured representation that maps items to latent attributes or skills, enabling interpretable modeling of what an item is intended to measure. Competence in fields like manufacturing is rarely a single trait and is better represented as a profile of partially mastered concepts and reasoning skills [9].

Our prior work introduced the Manufacturing Adviser as a GPT-based corpus of answers to 4 categories of manufacturing questions (process, sub-process, parameters, and design-for-manufacturing) adaptively worded for beginners and experts. Building on the forgoing literature, we extend this foundation by utilizing reverse Q/A (diagnostic questioning) for estimating learner expertise and adapting subsequent interaction. We incorporate Bloom-aligned WH-questions with a Q-matrix to measure manufacturing knowledge level across increasingly demanding cognitive operations rather than treating leaners with a vague or incomplete label.

2. Methodology

2.1. Bloom's taxonomy as the conceptual basis for reverse Q&A

Diagnostic questioning deliberately asks learners questions designed to elicit evidence of understanding before and during instruction. The question items do not have a single difficulty dimension. Instead, they are staged across Bloom's revised taxonomy, which organizes learning objectives by the type of cognitive work a learner performs. In Bloom's revised taxonomy, the progression moves from recalling knowledge to constructing increasingly complex reasoning: Remember, Understand, Apply, Analyze, Evaluate, and Create. This progression can help distinguish learners who repeat manufacturing terms from those who can reason about process constraints, causal mechanisms, and trade-offs. Remember and Understand levels show stable conceptual anchors necessary for later reasoning. Learners Apply knowledge in a use scenario, which is often where manufacturing competence becomes observable. At the Analyze and Evaluate levels, the learner must explain cause & effect and justify trade-offs among competing objectives such as cost, quality, throughput, tolerance, and manufacturability. To Create, manufacturing expertise synthesizes a plausible plan, reflecting the integrative decision-making. The tutor can treat Bloom level as an "evidence intent" tag attached to each Q/A probe. A short definition may be acceptable evidence at the Understand level, but it is not acceptable evidence at Analyze or Evaluate unless the learner expresses causal structure, constraints, and defensible justification. Table 1 shows Bloom's stages and cognitive learning application with examples.

The Bloom-tagged prompts also interpret responses for adaptive tutoring. Long answers lacking specific terms indicate shallow understanding, while a learner who is partially correct but uses constraints and causal structure implies developing expertise. The tutor offers clarification and vocabulary support for missing facts, and guided explanation for weak reasoning links.

Table 1: Bloom stages with manufacturing-oriented examples of diagnostic questions

Bloom stage	What the tutor is trying to elicit	Example manufacturing diagnostic question
Remember	Accurate recall of basic facts and terms	What is the definition of casting, and what is a mold?
Understand	Conceptual explanation in learner's own words	Explain, in plain language, why casting can produce complex shapes more easily than machining.
Apply	Using knowledge to propose an action in a concrete situation	A casting shows shrinkage porosity. What process change would you try first, and what outcome do you expect?
Analyze	Cause–effect reasoning and identification of contributing factors	A machined surface shows chatter marks. Explain two plausible causes and how you distinguish them.
Evaluate	Trade-off reasoning and justification under constraints	For low-cost tight tolerance mid-volume batch, would you select machining, forming, or casting? Justify your choice.
Create	Synthesis of a feasible plan or workflow	Propose a full process plan for an aluminum bracket, including the rationale for the sequence of operations.

2.2. Q-matrix as the skill model that makes level estimation interpretable

Besides cognitive depth of evidence, an expertise-adaptive tutor should support diagnosis for targeted adaptation. A Q-matrix is used to map assessment items to latent skills. It allows interpreting a learner's performance as a skill profile. We use a binary Q-matrix for, useful in early-stage educational research. We combine Bloom's taxonomy—cognitive depth of the requested evidence–and the Q-matrix—which manufacturing skill attributes are being activated—for a more robust measurement, e.g., performing well on Remember/Understand questions mapped to simpler General Process concepts but poorly on Apply/Analyze questions mapped to more advanced topic of Process Parameters implies a higher conceptual level in one attribute and a developing level in another.

2.3. Assessment design and evaluation workflow

To evaluate the feasibility of this diagnostic in manufacturing education, we implemented 20 multiple-choice questions (MCQ), 5 items per skill category, reporting skill-wise scores and a total score for an overall expertise label: Novice 0–5, Developing 6–10, Advanced 11–15, Expert 16-20. We conducted a pilot study with 8 students to determine feasibility, testing outcomes interpretability, and if skill-level reporting reveals patterns that an overall score can hide. Lacking an open-ended probing conversation data, we used an LLM-based simulation to generate responses across different

Table 2. Binary Q-matrix in the MCQ screening stage (20 items × 4 skills)

Item block	General Process	Sub-process	Process Parameters	DFM
Q1–Q5	1	0	0	0
Q6–Q10	0	1	0	0
Q11–Q15	0	0	1	0
Q16–Q20	0	0	0	1

presumed expertise levels. For each manufacturing topic, responses to one question were simulated at 3 assigned roles: novice, developing and expert.
The intention of this step is to validate whether reverse Q/A can, in principle, elicit level-differentiating response characteristics such as vocabulary, specificity, and explicit cause–effect reasoning, before scaling to studies with human participants. Yet, MCQs signal an initial skill profile fast, and conversations reveal higher Bloom levels where manufacturing expertise is most visible through language.

3. Results

In the eight-student MCQ pilot, the overall scores alone were not sufficient to describe learner capability because they tended to mask skill-specific patterns, supporting the methodological argument for skill-level reporting rather than relying only on a total score.
The pilot also surfaced an important item-design implication. Some participants performed better on the design-for-manufacturing items than on general process items, which may reflect recent exposure to design-focused coursework, but may also indicate that some distractors were not sufficiently discriminating. From the perspective of the overall-level mapping, approximately half of the participants fell into the Developing category, which is consistent with the expectation for a cohort of mechanical engineering undergraduates and suggests the scale is not producing implausible placements at this early stage.
The LLM-based role simulation produced qualitatively different response styles aligned with the intended expertise strata. Novice level responses tended to be short and uncertain with minimal technical vocabulary, developing level responses were generally correct but incomplete, and expert level responses included correct terminology and explicit cause–effect and trade-off reasoning. This supports the claim that Bloom-staged, diagnostic questions probing can be used to elicit and evaluate differences in reasoning depth, which is the core mechanism required for an expertise-adaptive Manufacturing Adviser. At the same time, the current results remain exploratory due to the small sample size and the use of simulated conversational data, which we identify as a limitation motivating the next phase of evaluation.

4. Conclusion

In this paper we proposed an expertise-adaptive framework for a Manufacturing Adviser that moves beyond reactive question answering by treating diagnostic questioning as a tutoring mechanism. The central argument is that a learner's manufacturing knowledge level becomes measurable and actionable when questions are intentionally designed to elicit evidence at progressively deeper cognitive levels and when responses are interpreted through an explicit skill structure. Bloom's revised taxonomy provides the pedagogical logic for staging diagnostic questions so that the tutor can distinguish between learners who can recall and describe manufacturing concepts and learners who can apply those concepts, explain causal mechanisms, and justify trade-offs under realistic constraints. The Q-matrix provides the complementary measurement structure that makes the inference interpretable at the skill level, enabling the tutor to represent manufacturing competence as a profile rather than a single total score.

The initial evidence supports the motivation for this design. The pilot MCQ diagnostic demonstrated that a small set of carefully categorized questions can generate skill-specific patterns that would be obscured by an overall score alone. The open-ended probing, although explored through simulated responses, further illustrates why reverse Q/A can be essential for an adaptive manufacturing tutor. Explanatory responses reveal reasoning depth, specificity, and constraint awareness that are difficult to observe in MCQ. Together, the results suggest that a Bloom-staged diagnostic dialogue coupled with Q-matrix-based skill mapping can estimate a learner's manufacturing knowledge level and drive pedagogically practical tutoring adaptations.

The next phase of this research will focus on strengthening validity, expanding coverage, and operationalizing more nuanced adaptivity. The first step involves extending the Q-matrix beyond a binary format to incorporate partial correctness, mixed evidence, and multi-skill integration through ordinal or weighted entries that reflect degrees of reliance and mastery. The second step underlies evaluating the reliability of open-ended conversations in education which requires constructing a clear rubric with defined scoring anchors, validating the rubric against human raters, and quantifying inter-rater agreement. The third step is to expand the empirical evidence base through larger and more diverse human studies. Future evaluations will increase sample size, capture more varied learner backgrounds, and track whether the inferred skill profiles correlate with independent measures such as course performance or instructor assessment. Beyond correlational validation, the central educational question is whether adaptivity improves learning efficiency and conceptual transfer. This requires study designs that compare non-adaptive and adaptive tutoring conditions and that measure not only correctness but also gains in explanation quality and reasoning depth, which are central to manufacturing expertise. Finally, the tutoring policy itself will be expanded from coarse adaptation toward more precise instructional decision-making. Future versions will tie adaptation explicitly to the combined Bloom–Q-matrix representation, so that the Adviser can diagnose not only which manufacturing skill is weak, but at what cognitive level the weakness appears. For example, a learner may understand a parameter concept but fail to apply it in a novel scenario; the tutor's response in that case should differ from the response to a learner who cannot define the concept at all. Implementing this distinction will require a larger question bank, broader coverage across manufacturing processes, and scenario-based cases that allow learners to demonstrate transfer across contexts—all can be potentially simulated with LLMs. These extensions will move the Manufacturing Adviser closer to a semi-autonomous tutor that can both diagnose and teach manufacturing engineering concepts in a structured, transparent, and scalable way.

Acknowledgment

This work is supported by the National Science Foundation under Grant No. 2229260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] D. Ismael and V. Jovanovic, "Advancing manufacturing engineering education through sustainability integration," *Manuf. Lett.*, vol. 46, pp. 37–41, Dec. 2025, doi: 10.1016/j.mfglet.2025.10.008.
[2] A. Sethupathy, P. Thakkar Singh, G. Gooding, J. Reich, D. Soukup, M. Lamarre, and J. S. Warrick, *Industry 4.0 and Modernizing Manufacturing Education*, Phase 1: Research and Documentation Report, American Society of Mechanical Engineers (ASME), New York, NY, USA, Feb. 2024

[3] J. Grodotzki, B. T. Müller, and A. E. Tekkaya, “Introducing a general-purpose augmented reality platform for the use in engineering education,” *Adv. Ind. Manuf. Eng.*, vol. 6, p. 100116, May 2023, doi: 10.1016/j.aime.2023.100116.
[4] F. Caccavale, C. L. Gargalo, K. V. Gernaey, and U. Krühne, “Towards Education 4.0: The role of Large Language Models as virtual tutors in chemical engineering,” *Educ. Chem. Eng.*, vol. 49, pp. 1–11, Oct. 2024, doi: 10.1016/j.ece.2024.07.002.
[5] M. Giannakos *et al.*, “The promise and challenges of generative AI in education,” *Behav. Inf. Technol.*, vol. 44, no. 11, pp. 2518–2544, July 2025, doi: 10.1080/0144929X.2024.2394886.
[6] W. Villegas-Ch, D. Buenano-Fernandez, A. M. Navarro, and A. Mera-Navarrete, “Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback,” *Smart Learn. Environ.*, vol. 12, no. 1, p. 41, June 2025, doi: 10.1186/s40561-025-00389-y.
[7] S.J. Hsieh and Q. Li, “Lessons Learned from an Intelligent Tutoring System for Computer Numerical Control Programming (CNC Tutor),” in *2018 ASEE Annual Conference & Exposition Proceedings*, Salt Lake City, Utah: ASEE Conferences, June 2018, p. 30762. doi: 10.18260/1-2--30762.
[8] J. Jackson, “Higher order prompting: Applying Bloom’s revised taxonomy to the use of large language models in higher education,” *Stud. Technol. Enhanc. Learn.*, vol. 4, no. 1, Feb. 2025, doi: 10.21428/8c225f6e.0915c17e.
[9] Y. S. Lim, “Gauging Q-Matrix Design and Model Selection in Applied Cognitive Diagnosis,” *Appl. Meas. Educ.*, vol. 37, no. 4, pp. 412–429, Oct. 2024, doi: 10.1080/08957347.2024.2438968.