https://doi.org/10.37420/j.emr.2026.018

# Human–AI Collaboration and Professional Judgment in Accounting Education

**Huang Yewei ***
**School of Accounting, College of Modern Economics and Management of Jiangxi University of Finance and Economics, Gongqingcheng, Jiangxi, China**
***Corresponding author, E-mail: 501297333@qq.com**

## Abstract
*Artificial intelligence can help accounting students produce analyses that exceed what they could complete independently. This creates a curriculum problem: better work produced with assistance does not necessarily indicate better understanding. This paper examines how human–AI collaboration can support the development of professional judgment rather than replace the reasoning through which it is learned. Drawing on cognitive offloading, research on human reliance on automation, and theories of curriculum alignment and feedback, it develops a conceptual framework connecting student learning with the improvement of teaching resources. The analysis identifies three tensions: greater access to explanations increases the need to evaluate them; automated assistance improves immediate performance while potentially removing opportunities to learn; and extensive learning records do not necessarily provide valid evidence of competence. Four mechanisms address these tensions: allocating cognitive tasks according to learning objectives, requiring verification against evidence, adjusting support as competence develops, and converting reviewed classroom feedback into improvements in cases and guidance. The framework distinguishes the learner's revision of a judgment from the institution's revision of teaching resources and assigns responsibilities to students, teachers, and AI systems. A management accounting example illustrates how these mechanisms operate together. The paper explains why the educational effects of AI depend on prior knowledge, assessment incentives, and the quality of human review, and proposes ways to examine both independent performance and responsible use of assistance.*

## Keywords
*accounting education; human–AI collaboration; professional judgment; cognitive offloading; curriculum reform; formative assessment*

## 1. Introduction

Artificial intelligence changes what an accounting student can produce before it necessarily changes what the student understands. A learner can obtain an explanation of a financial ratio, generate a spreadsheet formula, or draft a recommendation without working through the underlying reasoning. The output may be useful and even correct. Yet a teacher who assesses only that output cannot readily determine whether the learner understands the assumptions, could recognize an error, or would reach a defensible conclusion when the cir-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

cumstances change. The central problem for curriculum reform is therefore the relationship between assisted performance and the development of professional competence.

This distinction matters because accounting work combines calculation with interpretation and responsibility. The arithmetic of an investment appraisal can be automated, but the relevance of a cash flow depends on the decision being considered. A financial reporting estimate may draw on large quantities of information, but its justification still depends on the quality and applicability of that information. When AI contributes an explanation as well as a calculation, the student must evaluate not only a result but also the reasoning offered in support of it. Access to an explanation does not establish the capacity to judge whether it is appropriate.

Research on AI in accounting education has accordingly raised questions about the purpose of the curriculum. Ballantine, Boyce, and Stoner call for renewed attention to the human and critical dimensions of accounting rather than an exclusively technical response to AI [1]. In Chinese accounting education, Long proposes a move toward reciprocal relationships in which intelligent systems support teaching and classroom experience informs their improvement [2]. This broadens the discussion from isolated applications to the organization of teaching. It also leaves a question requiring closer analysis: which forms of interaction develop the learner's judgment, and which merely allow more work to be completed with assistance?

The present paper addresses that question by distinguishing learning from task accomplishment. In a workplace, delegating a familiar calculation can free an experienced accountant to consider a more difficult decision. In a classroom, delegating the same calculation may remove a novice's opportunity to understand a relationship that later judgments require. The educational consequences of delegation therefore depend on what the learner already knows and what the task is intended to teach. A practice that increases productivity cannot be assumed to develop competence for every student.

A second distinction concerns feedback. AI-generated advice can be immediate and detailed, but its educational value depends on whether students can evaluate and use it. A fluent explanation may encourage acceptance before its evidential basis is examined. At the institutional level, recording student interactions can help identify difficulties, but the records do not interpret themselves. An incorrect response might arise from a conceptual misunderstanding, an ambiguous instruction, an unreliable source, or a misleading system answer. Improving the course requires distinguishing these causes rather than treating all errors as equivalent training material.

This paper develops a conceptual account of how human–AI collaboration can support professional judgment in undergraduate accounting education. It connects research on cognitive offloading and reliance on automation with curriculum alignment and formative assessment. The argument is based on a problem-oriented synthesis of selected literature; it does not report a systematic review or a completed classroom experiment. The accounting situations used below are hypothetical illustrations. They make the proposed relationships explicit without presenting anticipated benefits as observed results.

The contribution is threefold. First, the paper explains why gains in assisted task performance can coexist with weak independent understanding. Second, it identifies mechanisms through which the allocation, review, and gradual withdrawal of assistance may convert collaboration into learning. Third, it distinguishes two processes of improvement: students revise their judgments, while teaching teams use reviewed evidence from

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

those revisions to improve cases, instructions, and reference materials. Connecting these processes provides a basis for sustained curriculum reform without assuming that interaction automatically improves an AI model.

## 2. Theoretical foundations and the meaning of collaboration

### *2.1. Professional judgment as the educational objective*

Professional judgment is the reasoned selection and justification of an accounting interpretation or course of action when rules, information, or objectives do not uniquely determine an answer. It requires sufficient technical knowledge to identify the issue, evidence to evaluate possible explanations, and awareness of the consequences and limits of a decision. This definition does not imply that all accounting questions are open to personal preference. A judgment must remain consistent with the relevant accounting requirements and the facts that can be established.

For teaching purposes, competence has both an independent and a collaborative dimension. Independent competence concerns what students can explain and apply without receiving the answer from another source. Collaborative competence concerns their ability to use assistance critically, including recognizing when to seek help and how to check it. These dimensions are related but not interchangeable. A student may perform well without tools yet be unable to detect a misleading generated explanation. Another may obtain a correct assisted answer without understanding why it is correct. A curriculum must provide evidence about both.

Human–AI collaboration is used here in a limited, operational sense: teachers and students use AI outputs during teaching and learning while retaining responsibility for educational goals, evaluation of evidence, and final decisions. It does not imply equal authority between humans and a system, nor does it attribute professional accountability to the system. The relevant division of work concerns particular cognitive activities, such as searching, calculating, comparing, interpreting, and reviewing. The distribution of those activities can change as students learn.

### *2.2. Cognitive offloading and the difference between performance and learning*

Risko and Gilbert describe cognitive offloading as the use of external actions or resources to reduce the cognitive demands of a task [3]. Their review examines both the decisions that lead people to offload and the consequences of doing so. This perspective helps explain why assistance is educationally ambiguous. External support may make a task manageable, but reducing the effort required for completion does not necessarily produce the understanding that the task was designed to develop.

The argument developed here distinguishes delegating work that obstructs learning from delegating work through which learning occurs. Automating repetitive formatting may allow students more time to compare alternative cost assumptions. Automating the selection of those assumptions may remove the central learning activity. The distinction is specific to the course and the learner: checking data joins might be incidental in one advanced decision course and essential in an introductory analytics course. A fixed rule that all routine work should be automated would overlook this dependence on educational purpose.

The wider literature also challenges a simple separation between automation and human enhancement.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Raisch and Krakowski analyze their interdependence in management [4]. Extending this concern to education requires an additional consideration: learners are acquiring the capabilities on which effective collaboration later depends. They cannot always review a delegated decision using competence already established. The educational design must therefore consider how today's allocation of work changes the student's capacity to perform and supervise related work tomorrow.

### *2.3. Reliance on AI and the role of evidence*

Parasuraman and Riley distinguish appropriate use of automation from misuse, disuse, and abuse [5]. Their analysis directs attention to how people rely on a system, rather than treating access to automation as inherently beneficial. In accounting teaching, both uncritical acceptance and blanket rejection of AI can be limiting. Students need to distinguish outputs that are sufficiently supported for the task from outputs requiring correction, qualification, or further investigation.

The paper refers to this ability as verification: checking a claim against sources, calculations, assumptions, and relevant circumstances. Verification is more demanding than asking a system whether its answer is correct. A second answer from the same system may repeat the same error. Independent checks could include tracing a number to a record, recalculating a material amount, or consulting the version of an accounting requirement specified for the case. The appropriate check depends on the nature and consequence of the claim.

AI-generated explanations also change the sequence of reasoning. An early answer can frame what students search for next and which alternatives they consider. A learner who first accepts that a cost is irrelevant may look only for support for that classification. Asking students to identify the decision and state an initial rationale before consulting a system makes the subsequent comparison more informative. This is a proposed instructional response to the risk of premature acceptance, not a claim that one sequence will suit every task.

### *2.4. Curriculum alignment and feedback as complementary explanations*

Constructive alignment links intended learning outcomes with teaching activities and assessment [6]. In an AI-supported course, this requires more than adding digital skills to an outcome statement. If the intended capability is to evaluate assumptions, students must undertake that evaluation during learning and demonstrate it in assessment. A course becomes misaligned when the system performs the central judgment but the final output is treated as evidence of the student's mastery. The discrepancy concerns who performed the relevant intellectual work.

Feedback theory explains how students can improve the work that remains their responsibility. Nicol and Macfarlane-Dick emphasize self-regulated learning, while Carless and Boud explain the importance of learners making sense of feedback and acting on it [7,8]. These perspectives support a distinction between receiving a correction and understanding its implications. Boud and Falchikov's concern with longer-term learning further directs attention to whether students can apply the resulting understanding beyond the original task [9].

Taken together, the theories explain different stages of the proposed process. Cognitive offloading identifies what is delegated. Research on automation identifies problems of reliance. Alignment asks whether the remaining activity serves the learning outcome. Feedback explains how that activity can become more com-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

petent through review and revision. Their integration does not imply that all assistance is undesirable. It specifies the conditions under which assistance may free students to reason, help them learn to reason, or replace reasoning they have not yet learned to perform.

## 3. Structural tensions in AI-supported accounting education

### *3.1. More explanations increase the need for discrimination*

When generating an explanation becomes inexpensive, the educational difficulty shifts toward deciding which explanation deserves reliance. Several plausible accounts of a falling margin may be available, but only some may fit the records. A system can describe product mix, input prices, and capacity utilization without establishing which actually matters. Students therefore need to distinguish a possible explanation from one supported by evidence. Producing a longer discussion does not resolve this distinction.

The problem is particularly clear in accounting because similar words can refer to different decision contexts. A cost allocation useful for reporting may be unsuitable for a short-term operational choice. A treatment that applies under one set of facts may not apply when contractual conditions change. Generated advice that omits these qualifications can remain persuasive. Teaching must make the relationship between the rule, its conditions, and the available evidence visible enough for students to examine.

A course organized mainly around obtaining answers leaves this problem unresolved. Students may be rewarded for collecting explanations without comparing their support. Reform should instead make the selection of evidence part of the task: which fact distinguishes the alternatives, where could that fact be found, and what conclusion is justified if it remains unknown? This changes the learning demand from assembling plausible statements to developing a defensible account of the problem.

### *3.2. Better immediate performance can conceal weaker development*

Assistance can improve the quality of a submission while reducing the learner's engagement with the reasoning that produces it. This is not inevitable, but it is a plausible consequence when the system supplies both the method and the interpretation. The student may learn how to obtain a satisfactory response without learning how to evaluate it. Immediate performance and later independent performance can then diverge, making the final assignment an incomplete indicator of educational success.

The divergence depends partly on prior knowledge. A well-prepared student can use an automated calculation to explore alternatives and recognize inconsistent results. A novice may not know which assumptions require examination. Giving both students identical assistance can therefore produce different consequences. At the same time, withholding all support from the novice may make the task inaccessible. The design problem is to provide enough support for participation while preserving opportunities to perform the reasoning that the learner needs to acquire.

This creates a temporal tension in curriculum reform. Activities that maximize completion today may not maximize capability tomorrow. Teachers need to decide which difficulties are productive and which merely consume time. Difficulty arising from a poorly organized file can often be reduced without educational loss.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Difficulty arising from comparing two defensible interpretations may be central to learning. The appropriate use of AI depends on understanding that difference rather than treating all saved effort as an equal benefit.

### *3.3. More learning data do not guarantee better educational decisions*

Digital systems can preserve submissions, questions, feedback, and revisions. Such records may support closer examination of learning, but their availability does not establish the validity of the inference drawn from them. Time spent on a task can indicate engagement, confusion, or an interruption. Repeated questions can reflect persistence rather than low ability. A polished explanation can indicate understanding, extensive assistance, or both. These ambiguities matter when records are used to allocate support or award grades.

The same limitation applies to resource improvement. Repeated student errors do not by themselves show that the reference explanation is correct or that the model needs further training. The cause may be an ambiguous scenario, a missing document, or inappropriate assessment criteria. If these distinctions are ignored, a system can reproduce the teacher's mistaken assumption more consistently rather than improve instruction. Human review is therefore part of the mechanism of improvement, not an administrative step added after data collection.

These three tensions connect the quality of information, the distribution of cognitive work, and the interpretation of educational evidence. Figure 1 summarizes their implications. The argument is that curriculum reform must respond to all three: reliable content cannot compensate for tasks that bypass learning, and demanding tasks cannot compensate for assessment that confuses assisted completion with competence. The relationships establish the need for a coordinated teaching design rather than an accumulation of separate digital functions.

| What AI makes easier | | What remains difficult | | Curriculum response |
|---|---|---|---|---|
| Generate explanations | → | Judge which claims are supported | → | Require evidence and comparison |
| Complete assignments | → | Distinguish assistance from understanding | → | Vary support and test new applications |
| Record learning activity | → | Identify the cause of recurring errors | → | Review evidence before changing resources |

*Table 1. Tensions in AI-supported accounting education and corresponding curriculum responses.*

## 4. Mechanisms through which collaboration can support learning

### *4.1. Allocating cognitive work according to the learning objective*

The first mechanism concerns who performs the decisions that matter for learning. A teacher begins by

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

identifying the accounting concept to be developed and the decisions that require its use. Assistance can then be assigned around those decisions. In a relevant-cost exercise, a system might organize transaction descriptions or perform calculations, while students identify which costs change between alternatives. In a data-quality exercise, students would need to examine the transformations themselves. Delegation follows the purpose of the task.

This approach separates access to assistance from authority over the conclusion. Students can consult a system while remaining responsible for deciding whether its advice fits the case. Teachers retain authority over the learning objective and the standard of acceptable justification. The system contributes provisional calculations, explanations, or questions that can be inspected. Such a division allows collaboration without treating a generated recommendation as the default answer against which students are judged.

The educational effect depends on what students do with time or effort released by assistance. If automated calculations are followed immediately by submission, little additional reasoning may occur. If the saved time is used to compare assumptions or examine a conflicting record, assistance can extend the inquiry. The mechanism therefore requires an explicit subsequent task. Efficiency becomes educationally useful when it enables additional or better reasoning that the curriculum actually requires.

Proposition 1 states that AI assistance is more likely to improve subsequent independent judgment when it reduces demands outside the focal learning objective and students use the released capacity to examine consequential decisions. The benefit should be weaker when the delegated work contains the reasoning being assessed. This proposition predicts different outcomes for the same tool across courses and learner groups, rather than a general advantage for higher levels of AI use.

### *4.2. Requiring evidence before accepting or revising a conclusion*

The second mechanism concerns verification. A learner encounters a provisional answer, identifies the assumptions on which it depends, and compares those assumptions with the records. A disagreement becomes educationally useful when the student can investigate it. Without access to suitable evidence, asking students to criticize AI may produce guesses or superficial objections. Reliable reference material and a clear accounting question are necessary conditions for meaningful review.

Feedback should point to the reasoning that needs examination. In a special-order decision, “reconsider capacity” is more useful when students can inspect a production schedule and explain whether the order displaces other sales. The teacher may ask what fact would change the recommendation rather than supply the preferred conclusion. Students then revise the analysis or explain why the challenge does not alter it. Both responses can demonstrate learning if the reasoning is adequately supported.

This process also gives disagreement an explicit status. A system answer is not automatically wrong because a student questions it, and a student answer is not automatically wrong because it differs from a reference response. The relevant question is which interpretation is supported under the stated facts. Teachers should make clear which matters are settled by the case and which remain uncertain. Otherwise, apparent openness can become a guessing exercise in which students try to infer the instructor's hidden preference.

Proposition 2 states that reviewing AI-generated advice is more likely to strengthen professional judgment

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

when students can test material claims against suitable evidence and explain the resulting decision. Repeated interaction without independent verification may improve the fluency of a response while leaving understanding unchanged. The expected learning mechanism is the examination of a discrepancy, followed by a reasoned resolution, rather than the volume of dialogue itself.

### *4.3. Adjusting support as students become more competent*

The third mechanism concerns the development of independence over time. Support that allows a novice to begin a task may later become unnecessary or restrictive. A worked example can make the relation between a concept and a procedure visible. Continued dependence on the same example can make it difficult to determine whether the learner can recognize the issue in a different setting. Teachers therefore need evidence about when to reduce prompts, provide fewer completed steps, or require an initial response before assistance becomes available.

This adjustment should be based on performance rather than a fixed assumption that every student progresses at the same rate. One learner may understand relevant costs but struggle with file organization; another may operate the spreadsheet accurately while misunderstanding opportunity cost. They need different assistance. A short explanation of a decision is often more informative for this purpose than a general confidence rating or a completion score. Support can then address the particular barrier without taking over the whole task.

Assessment has a direct role in this process. An assisted assignment shows how a student works with available resources; a later task with changed facts examines whether the underlying concept has been understood. A brief discussion can ask the learner to explain what would happen if an assumption were reversed. The aim is not to prohibit authentic tool use, but to identify which competence belongs to the student and which performance depends on current assistance. Both are relevant to professional preparation.

Proposition 3 states that guided collaboration is more likely to develop transferable judgment when support is reduced in response to demonstrated understanding and students subsequently apply the concept to changed circumstances. Support withdrawn too early may increase errors without promoting learning; support retained after it is needed may conceal dependence. This predicts that the timing and content of assistance matter more than a uniform preference for either unrestricted access or complete exclusion.

### *4.4. Converting classroom feedback into reliable teaching resources*

The fourth mechanism operates beyond the individual learner. Reviewed classroom experience can reveal why a task succeeds or fails: an instruction may invite a mistaken interpretation, a reference answer may overlook a legitimate alternative, or an AI explanation may repeatedly confuse two concepts. Teaching teams can use these findings to revise the case and the guidance provided with it. The student's learning process thus supplies evidence for a separate process of curriculum improvement.

This second process requires a decision about what should change. A confirmed error in a reference document calls for correction of that document. A recurring misconception may justify a new comparison exercise. An unclear prompt may need rewriting. An unreliable model response may require restricting a use case

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

or adding a check. These are different interventions. They should not be described collectively as automatic model learning, because changing a document repository, changing instructions, and changing model parameters involve different processes and responsibilities.

Figure 2 distinguishes the two connected processes. Within a learning activity, the student develops a provisional judgment, examines feedback, and produces a justified revision. Across teaching cycles, reviewed difficulties inform changes to resources and subsequent tasks. Teachers mediate the connection by checking the meaning of the evidence and authorizing changes. A revised resource should be evaluated on later cases, including cases unlike those that prompted the revision, before it is treated as an improvement.

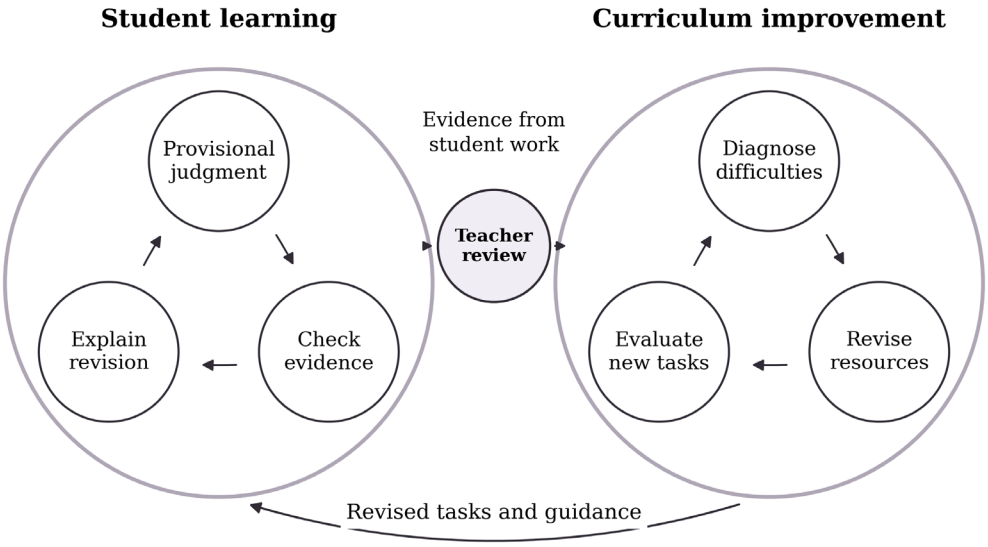



*Figure 2. Connecting student learning with the reviewed improvement of teaching resources.*

Proposition 4 states that classroom interaction is more likely to support sustained curriculum improvement when identified problems are reviewed, linked to a specific cause, and followed by documented and evaluated changes. Collecting more interaction data without these steps may increase the volume of records without improving teaching. The mechanism is institutional learning through checked evidence, not the assumption that accumulated data necessarily produce a more capable system.

Together, the four mechanisms connect decisions within a task with changes across teaching cycles. They also identify different failure points. A well-organized resource collection cannot compensate for delegating the concept students need to learn. Student reflection cannot correct a misleading source that no one checks. Carefully reduced support can still be undermined by assessment that rewards only the final product. Effective reform depends on the consistency of these arrangements, while evaluation must distinguish their separate contributions.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

# 5. Implications for accounting curriculum reform

## *5.1. Connecting the mechanisms through a management accounting case*

A special-order decision illustrates how the mechanisms operate together. Consider a hypothetical manufacturer offered an order for 1,000 units at 90 currency units each. Incremental production cost is 65 per unit, and the current costing report allocates fixed overhead of 30 per unit. A response based only on full unit cost would reject the order because 90 is below 95. A response based only on incremental cost would recommend acceptance because the order appears to contribute 25,000. Neither recommendation is adequately supported until the learner establishes whether capacity is available and whether the allocated overhead changes with the decision.

The teaching objective is the identification of relevant costs, not the speed of multiplication. Students first state a provisional recommendation and identify information they need. AI may calculate the amounts and offer competing explanations, but students must justify which costs belong in the comparison. The teacher then supplies a production schedule showing that the order would displace other business contributing 20 per unit of the same capacity. On these simplified assumptions, the net incremental benefit is 5,000, rather than 25,000. If the displaced contribution were instead 28 per unit, the net benefit would be negative 3,000. The changed fact alters the decision without changing the basic arithmetic method.

The educational value lies in explaining the change. Students should distinguish an allocation that remains unchanged from an opportunity cost that arises because resources have an alternative use. They should also identify assumptions omitted from the numerical comparison, such as price effects on existing customers and the reliability of the demand forecast. Those issues can justify a conditional recommendation, but they should not be assigned invented numerical values merely to complete the calculation. The distinction between known amounts and unresolved considerations is part of responsible judgment.

Feedback asks students to revisit the assumption that determined their initial answer. A learner who accepted the order should explain whether and why the capacity information changes the recommendation. A learner who rejected it on full-cost grounds should reconsider the relevance of the allocated overhead. A later individual task can use a make-or-buy decision with avoidable fixed costs, requiring the same principle in a different setting. This provides evidence about whether students have learned the concept or merely remembered the special-order answer.

At course level, the teacher reviews recurring errors before changing the resources. If students consistently treat allocated overhead as avoidable, a comparison example may be needed. If the production schedule is ambiguous, the case itself needs revision. If an AI answer assumes idle capacity without evidence, the guidance should require an explicit capacity check. These responses illustrate how the same classroom activity can support student learning and resource improvement, provided the cause of the difficulty is established rather than inferred from the final score alone.

## *5.2. Structuring inquiry without overprescribing the answer*

Teaching can introduce independence gradually through changes in the decisions assigned to students.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Research on simulation and instructional support provides a basis for attending to prior knowledge and guidance, although it does not establish the effects of the particular sequence proposed here [10]. An initial exercise may provide a worked example and ask why each step is necessary. A subsequent exercise can omit the method, requiring students to select and justify it. A later activity can introduce conflicting information or changing assumptions. The progression concerns responsibility for reasoning, not simply increasing the number of records or the complexity of the software.

Discussion should focus on decisions that differentiate stronger and weaker responses. In the special-order example, students might disagree about whether the production schedule establishes a genuine capacity constraint or whether the alternative sales are sufficiently certain to justify an opportunity cost. The teacher can ask what additional information would resolve the disagreement. This shifts attention from guessing the preferred answer to identifying the basis for a defensible conclusion. It also helps students recognize when a disagreement is about evidence and when it concerns the application of an accounting concept.

Collaboration can make these differences visible, but group work needs an educational purpose beyond dividing the workload. Students can review one another's assumptions, compare methods, or argue from different managerial perspectives. Roles should change so that no student is consistently responsible only for data entry or presentation. Individual explanations remain necessary because an excellent group report may conceal unequal understanding. Teachers should be able to identify what each learner can explain independently.

Generative AI can be included as an object of scrutiny within this sequence. Students may compare an explanation produced by a tool with the underlying records, identify unsupported claims, or test whether a recommendation changes when an assumption is altered. The learning demand lies in verification and correction. A requirement to submit a long record of prompts is not sufficient evidence of understanding. A concise explanation of what was accepted, rejected, and checked provides a more direct basis for evaluating the student's contribution.

### *5.3. Designing assessment around evidence of learning*

Authentic assessment connects educational tasks and criteria with relevant demands of practice [11]. In the present framework, this means that assessment criteria should describe the quality of reasoning in terms students can apply. Technical accuracy concerns whether records and calculations are reliable. Use of evidence concerns whether the information supports the claim. Interpretation concerns whether the student explains the relationship between the result and the business issue. Professional responsibility concerns the treatment of uncertainty, affected parties, and limits of the recommendation. These criteria should be illustrated with examples rather than left as broad labels on a marking sheet.

A course can collect evidence at a few meaningful points. An initial decision memo records the proposed method and assumptions. A revised analysis shows how the student responds to a substantive challenge. A final individual task uses unfamiliar records to examine whether the same principles can be applied independently. The exact weighting depends on course objectives and institutional requirements; a universal allocation would imply a precision that the conceptual argument does not support. What matters is that each component has a distinct evidential purpose.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Teacher moderation is particularly important when several conclusions may be defensible. Assessors can compare sample responses and discuss which differences reflect stronger reasoning rather than personal preference. The criteria should distinguish a different but supported judgment from a conclusion that overlooks relevant evidence. Where oral discussion forms part of assessment, comparable prompts and a clear recording method help maintain fairness. The assessment should test understanding without turning fluency in public speaking into a substitute for accounting competence.

Learning records can inform course improvement when interpreted alongside the work itself. Frequent incorrect classifications may indicate a conceptual problem; repeated data mismatches may indicate insufficient preparation; identical misunderstandings of a document may indicate ambiguous instructions. These diagnoses require different responses. A dashboard showing low completion cannot identify the appropriate remedy on its own. Teachers need to examine representative responses before changing content, support, or the assessment task.

### *5.4. Supporting teachers and maintaining curricular continuity*

Curriculum reform creates work that must be recognized. Developing a case with plausible alternatives, testing its records, preparing assessment criteria, and reviewing student reasoning all require time. Institutions should consider these demands when allocating teaching responsibilities. A resource that is technically impressive but too costly to maintain may be used inconsistently or abandoned. A smaller set of carefully developed tasks can provide a more sustainable basis for reform than a large collection of disconnected materials.

Collaboration with practitioners can strengthen the relevance of cases when responsibilities are clear. Practitioners may review business assumptions, identify realistic consequences, or discuss the feasibility of recommendations. Academic staff retain responsibility for learning objectives, suitable difficulty, and fair assessment. Professional realism and educational value overlap, but they are not identical. A task that is routine for an experienced accountant may overwhelm a novice, while a simplified case may illuminate a relationship that workplace complexity would obscure.

Across a program, teaching teams can agree where students first encounter particular forms of judgment and where they revisit them with greater independence. Financial accounting might introduce the justification of an estimate, management accounting might develop analysis of alternative actions, and a final integrative course might require a recommendation under several constraints. Coordination should make expectations intelligible without imposing one rigid teaching method. The common purpose is progression in responsibility for reasoning, supported by sufficient disciplinary knowledge.

## 6. Boundary conditions and directions for evaluation

### *6.1. Student preparation and equitable participation*

The framework assumes that students have, or can obtain, the foundational knowledge required for the assigned decisions. Where this condition is absent, open-ended tasks may produce confusion rather than useful inquiry. Teachers should distinguish a need for further explanation from a need for additional practice or technical assistance. Short diagnostic activities can guide this judgment. Support can then be reduced as students

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

become able to explain and perform the relevant work independently.

Participation also depends on access to reliable tools and usable materials. A task should not make success depend on a privately purchased application unless equivalent provision is available. Students may need alternative ways to inspect records, participate in discussion, or explain decisions. These arrangements should preserve the learning outcome while reducing barriers unrelated to it. Fairness requires attention to the demands actually placed on learners, rather than assuming that a common digital interface creates an equivalent educational opportunity.

### *6.2. Technology limits and institutional choices*

Technology should be selected according to the relationships that teaching requires. Useful capabilities may include access to source records, inspection of assumptions, comparison of alternatives, and preservation of selected revisions. These functions can be supported through different combinations of spreadsheets, learning management systems, document repositories, and specialist applications. No single architecture follows necessarily from the theoretical framework. The appropriate choice depends on course scale, existing provision, staff expertise, and maintenance costs.

Institutional decisions should also distinguish educational evidence from excessive monitoring. Recording every action can create large quantities of data without clarifying how students reason. A selective record of important choices is usually more closely connected to the assessment purpose. Where student work is later used for research, that use should be addressed through appropriate consent and institutional procedures. The teaching requirement to submit work and the research decision to analyze identifiable student records should remain distinguishable.

### *6.3. Evaluating learning and resource improvement separately*

Evaluation should distinguish immediate assisted performance, later independent performance, and the quality of decisions about when to rely on AI. These outcomes can move in different directions. A rise in assignment scores accompanied by weaker performance on a comparable new task would challenge the claim that assistance improved learning. Conversely, sound independent understanding alongside poor detection of misleading system advice would indicate that collaborative competence still needs attention. Reporting only one outcome would conceal this distinction.

Proposition 1 can be examined by varying which parts of a task are delegated while keeping the accounting content and available time comparable. Prior knowledge should be assessed because the same delegation may release capacity for one student and bypass essential learning for another. Proposition 2 concerns the contribution of verification. Comparisons should examine whether students have access to independent evidence, whether they actually use it, and whether their revisions address the material discrepancy. Counting prompts or feedback messages would not identify the proposed mechanism.

For Proposition 3, the relevant comparison concerns the timing of support. Students can encounter comparable tasks with different degrees of prompting, followed by a new application with limited assistance. Performance should be evaluated through a common rubric addressing conceptual accuracy, evidence, and jus-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

tification. Where possible, reviewers should assess responses without knowing which teaching arrangement produced them. Changes in confidence can help interpret the experience but should remain separate from demonstrated competence.

Proposition 4 requires evidence across teaching cycles. A team could record the difficulty identified, the basis for its diagnosis, the authorized resource change, and the result when the revised task is used. Evaluation should include unfamiliar cases so that improvement is not defined solely as better performance on examples used to revise the resource. If a change resolves one ambiguity but introduces another, the record should preserve that result. Resource maintenance is a process of testing educational decisions, not simply increasing the size of a case collection.

A credible study would also account for changes in practice time, teacher attention, and assessment incentives. Where random allocation is impractical, baseline measures and comparison classes can improve interpretation, although they do not remove all threats to causal inference. Qualitative examination of students' explanations can help establish whether the expected reasoning occurred. The purpose is to test the mechanisms and their limits, rather than demonstrate the success of a technology chosen in advance.

## 7. Conclusion

Human–AI collaboration changes both the work students can accomplish and the reasoning they need to learn. The two do not develop automatically together. This paper has explained how better assisted performance can coexist with limited understanding when technology takes over the judgments that a course intends to teach. It has also identified conditions under which assistance can support deeper learning by making evidence easier to examine and creating opportunities to compare, revise, and defend decisions.

The framework connects four mechanisms: allocating cognitive work according to the learning objective, verifying claims against evidence, adjusting support as understanding develops, and using reviewed classroom experience to improve teaching resources. Their relationship extends across two processes. Students learn through the reconsideration of particular judgments; teaching teams learn by diagnosing recurring difficulties and evaluating changes to cases and guidance. The connection depends on human review and clear responsibility, not on an assumption that more interaction necessarily improves a model.

The argument contributes a conditional explanation of AI-supported accounting education rather than a claim of demonstrated effectiveness. It suggests that curriculum reform should be judged by the quality of students' independent reasoning and their responsible use of assistance, alongside the reliability of the teaching arrangements that support both. Empirical work is needed to establish how these relationships vary with prior knowledge, task complexity, and institutional resources. Those differences are central to the theory because they determine whether collaboration extends the learner's capabilities or merely improves the work produced on the learner's behalf.

## Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

## Funding

This research was supported by the Jiangxi Provincial Higher Education Teaching Reform Research Project (Grant No. JXJG-23-34-4).

## References

[1] Ballantine, J.; Boyce, G.; Stoner, G. A critical review of AI in accounting education: Threat and opportunity. Critical Perspectives on Accounting 2024, 99, 102711. https://doi.org/10.1016/j.cpa.2024.102711.

[2] Long, Y. Constructing a new ecosystem of autonomous intelligent accounting education in China: From tool assistance to human–machine symbiosis [in Chinese]. Friends of Accounting 2026, 22, 2–9.

[3] Risko, E.F.; Gilbert, S.J. Cognitive offloading. Trends in Cognitive Sciences 2016, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

[4] Raisch, S.; Krakowski, S. Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review 2021, 46(1), 192–210. https://doi.org/10.5465/amr.2018.0072.

[5] Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. Human Factors 1997, 39(2), 230–253. https://doi.org/10.1518/001872097778543886.

[6] Biggs, J. Enhancing teaching through constructive alignment. Higher Education 1996, 32, 347–364. https://doi.org/10.1007/BF00138871.

[7] Nicol, D.J.; Macfarlane-Dick, D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education 2006, 31(2), 199–218. https://doi.org/10.1080/03075070600572090.

[8] Carless, D.; Boud, D. The development of student feedback literacy: Enabling uptake of feedback. Assessment & Evaluation in Higher Education 2018, 43(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354.

[9] Boud, D.; Falchikov, N. Aligning assessment with long-term learning. Assessment & Evaluation in Higher Education 2006, 31(4), 399–413. https://doi.org/10.1080/02602930600679050.

[10] Chernikova, O.; Heitzmann, N.; Stadler, M.; Holzberger, D.; Seidel, T.; Fischer, F. Simulation-based learning in higher education: A meta-analysis. Review of Educational Research 2020, 90(4), 499–541. https://doi.org/10.3102/0034654320933544.

[11] Gulikers, J.T.M.; Bastiaens, T.J.; Kirschner, P.A. A five-dimensional framework for authentic assessment. Educational Technology Research and Development 2004, 52(3), 67–86. https://doi.org/10.1007/BF02504676.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)