Print ISSN: 3104-932X
Online ISSN: 3104-9338

A Generative AI–Supported Instructional Model for Statistical Modeling in Undergraduate Probability and Statistics

Wen Zhong[1,2*], Yan Ze Teh[3]

[1] Nanning Normal University, Nanning, China
[2] Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia
[3] Xiamen University Malaysia, Sepang, Malaysia

Received: Nov 2, 2026

Revised: Nov 8, 2026

Accepted: Nov 19, 2026

Published online: November 20, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 3, No. 2 (February 2027)

* Corresponding Author: Wen Zhong (wenzhong@nnnu.edu.cn)

Abstract. This instructional design article proposes a six-stage model for undergraduate probability and statistics. It asks how a modeling unit can connect calculations under a specified model with statistical decisions when data definitions change, and how instructors can adjust support using students' initial responses and explanations of revisions. For each stage, the model specifies what students submit, what instructors check, and when students revisit an earlier decision. Instructors use generative AI (GenAI) only to draft follow-up questions for their review. Students respond to those questions without interacting with GenAI. A worked example shows how students move from a Poisson probability calculation to an order-level comparison using constructed logistics complaint records. The example also shows how instructors use students' explanations to decide whether to reduce prompts or revisit an earlier stage. Classroom feasibility and effects on student learning remain to be evaluated.

Keywords: *Probability and statistics; Statistical modeling; Generative AI; Instructional scaffolding*

1. Introduction

In undergraduate mathematics education, probability and statistics courses give students opportunities to use data to answer questions about specific situations. Students need to explain what they are analyzing, how the data were generated, and what conclusions their calculations support. Guidelines for undergraduate statistics education emphasize using real data in relation to the context and the question [1]. In one comparison of introductory statistics assessments, scores were lower on a revised version that replaced realistic synthetic data with data from published studies in some items. The difference may also have reflected the particular contexts

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

of those items [2].

When tools can generate analytical steps, instructors need to determine from students' responses whether students assessed the conditions for using the method themselves. Discussions of course materials and code generation, together with examples of tools answering conceptual questions, illustrate the procedural support GenAI can provide [3,4]. The version used in early examples of conceptual questions did not support data uploads [4]. Later work examined how a GenAI tool analyzed constructed data. The tool performed some operations, but it checked the conditions for using the methods inconsistently [5].

Students who generate code from natural-language prompts still need to interpret the statistical meaning of the output. Natural-language code generation has been used in course activities and tutoring in introductory data science and data analytics instruction [6,7]. In a task-based interview study, five undergraduates repeatedly adjusted their prompts and revised their plots. Their judgments when interpreting the plots were also related to their prior statistical knowledge [8]. When reviewing results, instructors also need to examine how students explain the relationships among variables, conditions, and output.

The same incorrect conclusion can arise from a misidentified statistical unit, an error in assigning records, or an unsuitable method. Instructors therefore need student responses that reveal where the difficulty arose. An earlier statistics course project prohibited the use of GenAI [9]. A subsequent experimental design project allowed GenAI use, offered optional virtual populations, and included checkpoints and feedback during exploratory data analysis and drafting [10]. These checkpoints give instructors access to students' preliminary judgments before the final submission.

The classroom materials available to instructors determine which difficulties can be addressed promptly. Tasks that involve comparing outputs or collaborating provide code and explanations for review [11,12]. Simulated tutoring can use information about the current interaction, while a framework for assigning tools to cognitive tasks uses task demands to define tool roles [13,14].

Prior studies describe the procedural help GenAI can offer and the student work instructors can use to examine statistical reasoning. This instructional design article addresses two questions. First, how can a modeling unit in undergraduate probability and statistics connect calculations under a specified model with statistical decisions when data definitions change? Second, how can instructors use initial responses and explanations of revisions to decide when to continue or reduce support and when to revisit an earlier stage?

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

We propose a model that combines three task types, six stages of statistical modeling, and instructor feedback. The model specifies what students submit and how instructors use that work to choose a follow-up question or identify which earlier decision needs review. Instructors use GenAI only to draft follow-up questions about difficulties they have already identified. They review each draft before sharing a question. Students work from the task materials and respond to questions provided by the instructor. The logistics example uses constructed records and illustrative responses to show how the model can be applied. The article's contribution is a teaching design and a worked example. Its classroom feasibility, effects on learning, and the added value of GenAI-assisted drafting remain open empirical questions.

2. Design Rationale and Overall Structure

2.1. Classroom Stages Based on Statistical Inquiry

Students may discover only after completing a calculation that the statistical unit represented in the records differs from the one required by the question. In statistical inquiry, questions, data, and conclusions can be revisited. New questions arising during analysis can be addressed by returning to the data or the original question [15,16]. The six stages begin with defining the problem and preparing the data. Students then select a method, solve and check the model, interpret and revise, and reflect, summarize, and extend their work. Method selection and computational checks take place in separate stages. Students first explain why they chose the method and later report what their checks showed. In Material B, the contact intake date alone cannot identify the order completion cohort. Students return to data preparation and link the contact to the completed order using its waybill number.

Table 1. Course Content and Corresponding Modeling Tasks

Course content	Possible context	Student tasks
Discrete random variables and their distributions	Counting complaints at a service center	Define the event and observation window, explain the parameter, and calculate and interpret the probability
Sampling distributions	Repeated sampling of delivery times	Distinguish individual delivery times from sample means and explain sampling variability
Parameter estimation	Estimating mean delivery time	Describe the population and sample, interpret the estimate, and explain the limitations of the data
Hypothesis testing	Comparing delivery times before and after a service change	Specify what is being compared and describe the data structure, explain the assumptions of the method, and interpret the results

Table 1 links each course topic and a possible modeling context to the work students are expected to do. Complaint counts give students practice in defining events and interpreting

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

parameters. Repeated sampling of delivery times helps students distinguish individual observations from sample means.

2.2. Task Structure and Student Decisions

Foundational, extension, and open exploration tasks differ in which conditions are specified and which decisions are left to students. Foundational tasks give students a relatively well-specified model and its parameters. Students interpret the variables, events, and results. Extension tasks change the research question or the data conditions, requiring students to decide how to adjust the statistical unit, counting rules, and method. In open exploration tasks, students further define the problem, identify what data they need, or compare possible approaches. Table 2 illustrates this arrangement using complaint counts.

Table 2. Task Conditions, Student Decisions, and Prompts

Task	Information provided	Decisions for students	Response indicating a need for support	Initial support
Foundational task	Distribution, parameter, and observation window	Define the probability event and interpret the result	Writes exactly 3 complaints for at least 3 complaints	Ask about the complement of the event
Extension task	Two order completion cohorts and records of complaint-related contacts	Determine the statistical unit, the date to which records are assigned, and the follow-up window	Uses daily counts of incoming contacts to judge the risk of a complaint at the order level	Ask students to use the waybill number to identify the order completion cohort
Open exploration task	Possible research objectives and leads on available data	Specify how to record the data, justify the method, and define what conclusions the data can support	Identifies data fields that do not address the research objective	Check data requirements against the objective

Students can meet the common foundational requirements within a topic and then choose a subsequent task. Instructors offer help based on the difficulty students are currently encountering, and task choices do not determine permanent group membership. All task types require students to retain a statement of the problem and an interpretation of the results. Figure 1 shows how the six stages organize the work across task types.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

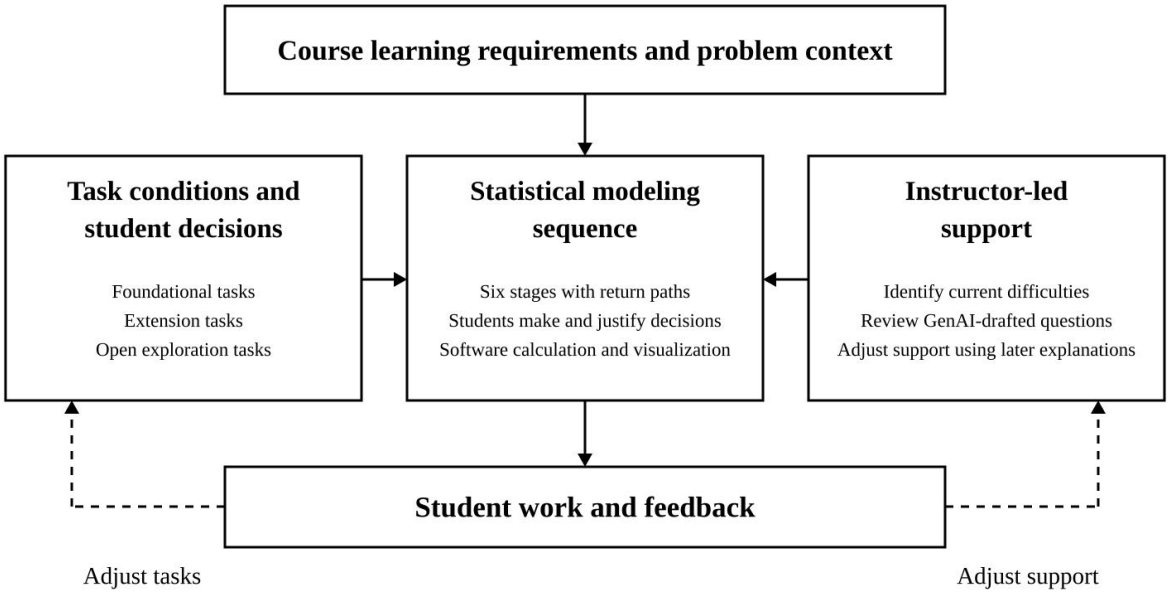


Figure 1. Overall Structure of the Instructional Model

Note. Feedback on students' work informs adjustments to both tasks and learning support.

2.3. Adjusting Support Based on Student Responses and Explanations

Once a student can correct an answer, the instructor must still decide whether to continue offering the same kind of help. Instructional scaffolding is adjusted to the learner's current difficulty and gradually withdrawn as the learner becomes able to make the relevant judgments [17]. The instructor first identifies the difficulty from the initial response and then responds with a follow-up question, material for comparison, or a demonstration of the step that caused difficulty. Once the student can explain the reason for the revision, the instructor temporarily reduces similar prompts and uses a subsequent task to check whether the student can answer independently. If the difficulty recurs under new conditions, the instructor restores the relevant support.

In the proposed model, the instructor uses GenAI solely to draft a follow-up question. The instructor first identifies the difficulty in the student's initial response. The instructor then gives GenAI a de-identified excerpt and asks for a question that addresses the difficulty. Before sharing the question, the instructor checks its mathematical wording and makes sure it leaves the statistical decision to the student. This instructor review is consistent with guidance on GenAI in education [18]. The instructor can use the same reviewed question with students whose responses show the same difficulty. Students revise their responses independently of GenAI and explain what they changed and why. They verify calculations with formulas or course software when appropriate.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

3. Six Instructional Stages and Adjustments to Support

Figure 2 shows the six stages and typical return paths. Each stage specifies a student submission, an instructor check, an anticipated difficulty, and the condition for revisiting earlier work. The instructor directs the return to the stage where the problem arose. Students recheck any later calculation or conclusion affected by the revision.

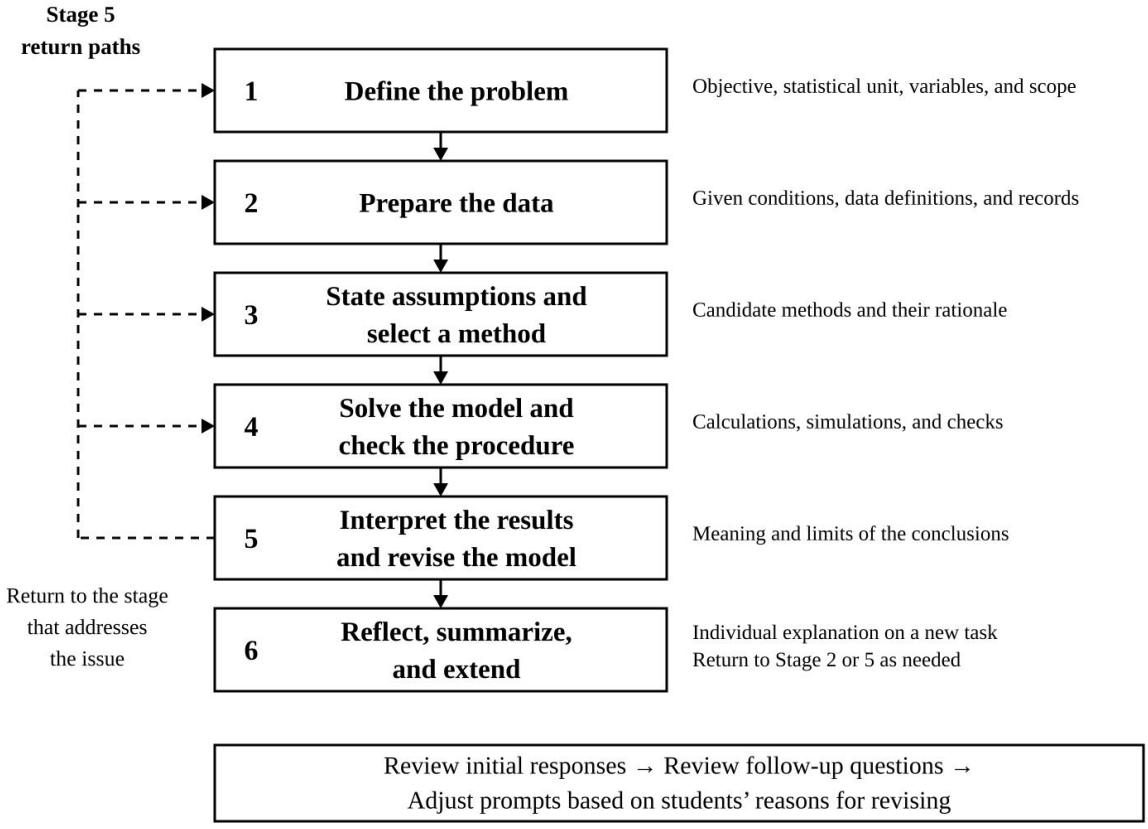


Figure 2. Six-Stage Instructional Sequence with Paths for Revision

Note. Dashed arrows show returns from Stage 5. Sections 3.2–3.6 specify the return conditions and the paths from other stages. In Stage 6, students return to Stage 2 to check record coverage or to Stage 5 to revise a conclusion.

3.1. Defining the Problem

Students submit a short problem statement that identifies the objective, the statistical unit, the variables, and the scope. The instructor checks whether the proposed unit and outcome match the objective. Students may confuse a question about contact workload with a question about the proportion of orders with a complaint. Students clarify the question at this stage. They return to Stage 1 from a later stage if the available information or intended comparison requires a change in the original objective.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

3.2. Preparing the Data

Students submit a data summary or a collection plan. They identify the fields needed to answer the question and document how they count events, assign records, and handle duplicates and missing information. They also specify the relevant observation or follow-up window. For a task with a specified distribution, they list the given conditions and the data they would need to check them in practice. The instructor checks whether the records support the proposed statistical unit and cover the required dates. Anticipated difficulties include treating contact intake dates as order completion dates, counting repeat contacts as separate orders, and treating an unobserved day as having zero complaints. Students correct assignment and coverage errors in Stage 2. They return to Stage 1 if the available records require students to redefine the objective or statistical unit.

3.3. Formulating Assumptions and Selecting a Method

Students submit their recommended method, define the variables and any parameters, and explain which task conditions or record features justify the method. The instructor checks whether these reasons support the proposed calculation or comparison. An anticipated difficulty is assuming that nonnegative integer counts are sufficient to justify a Poisson model. If the justification depends on unchecked recording conditions or missing data, students return to Stage 2. If the method choice exposes an ambiguity in the objective, students return to Stage 1 to clarify it. When the available evidence cannot support a distributional choice, students state what additional information they need.

3.4. Solving the Model and Checking the Procedure

Students submit the calculation or software output along with the inputs and a brief account of their checks. The instructor checks whether the event, inputs, and procedure match the chosen method. An anticipated difficulty is writing the event for at least three complaints as exactly three complaints. Students correct event notation and arithmetic in Stage 4. They return to Stage 3 if a check undermines an assumption of the selected method, or to Stage 2 if the inputs reveal an error in record assignment or coverage.

3.5. Interpreting Results and Revising the Model

Students submit a conclusion in context along with any revision and their reason for making it. The instructor checks whether the conclusion answers the original question and stays within the scope supported by the statistical unit, records, and method. An anticipated difficulty is

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

interpreting counts of incoming contacts as the proportion of orders with a complaint. Students revise an overstatement in Stage 5 if the underlying analysis is sound. They return to Stage 4 for a calculation error, Stage 3 for an unsupported method, Stage 2 for a data definition or coverage error, or Stage 1 if the objective itself needs to change.

3.6. Reflecting, Summarizing, and Extending

Each student submits an individual response to a new task with a changed condition. The response identifies a decision that should be kept or revised and explains why. The instructor checks whether the student can apply the relevant reasoning with fewer prompts. An anticipated difficulty is assuming that a comparison is possible when the records end before follow-up is complete. If the student treats an unobserved day as having zero complaints, the instructor directs the student back to Stage 2 to check the dates. If the student identifies the incomplete follow-up window but still claims that the full-window proportions can be compared, the instructor directs the student back to Stage 5. The instructor keeps prompts reduced for decisions the student can justify and provides targeted support for unresolved difficulties.

4. A Worked Teaching Example on Logistics Complaints

4.1. Scenario and Materials

The example is intended for undergraduates who have studied discrete random variables, the Poisson distribution, and complementary events. It includes two worked tasks involving a logistics service point, followed by an optional open exploration task. In the first worked task, students use a specified model to calculate the probability of at least three complaint events in one workday. In the second, they use records of complaint-related contacts to compare the proportions of completed orders with at least one recorded complaint within two full calendar days after completion. The second task ends with an individual check using a new set of records.

Material A sets out the assumptions for a mathematical exercise. The observational unit is one full workday, which also defines the observation window. The task applies a consistent rule to remove duplicate records and count distinct complaint events. Let N denote the number of these events in one workday. The task specifies that N follows a Poisson distribution with parameter 2.

Material B is a separate set of records constructed for instruction. The student responses and instructor questions are illustrative. The statistical unit is a completed order. Each raw contact record represents one complaint-related contact. Codes such as A01 and B01 are waybill

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

identifiers that link the contact records to the order records. On June 1, 4 orders (A01 through A04) were completed. On June 2, 8 orders (B01 through B08) were completed. The records extend through the end of June 5 and include all complaint-related contacts received for the listed orders through that date. One contact concerning A01 was received on June 2. On June 3, one contact was received for A01 and one for B01. Contacts concerning B02 and B03 were received on June 4 and June 5, respectively. The follow-up window for each order covers the two full calendar days immediately after its completion date, excluding the day of completion. Students assign each order to a cohort by its completion date and use contact intake dates to check whether each contact falls within the order's follow-up window. Material B is not used to estimate the parameter in Material A. Table 3 shows how the two tasks proceed and what the instructor uses to guide feedback.

4.2. Using Students' Event Definitions to Guide Computational Support

The foundational task asks students to explain what N represents and to write the event corresponding to at least 3 complaint events. The complement of the event "at least 3 complaints" consists of the counts 0, 1, and 2. The Poisson probability mass function gives the probability of each count [19]. Under the conditions specified in Material A, the probability is

$$P(N \geq 3) = 1 - P(N = 0) - P(N = 1) - P(N = 2)$$

$$= 1 - \exp(-2) \times (1 + 2 + 2^2/2!) = 1 - 5\exp(-2) \approx 0.3233.$$

Suppose a student writes the desired probability as $P(N = 3)$. After reviewing the response, the instructor can ask, "Which counts are included in fewer than 3 complaint events?" If the student lists 0, 1, and 2 and explains why these counts form the complement of the original event, the instructor can temporarily withdraw similar prompts about expressing the event and retain the computation check. If the student continues to confuse "at least" with "exactly," the instructor can briefly demonstrate how to identify the range of counts and then ask the student to rewrite the explanation.

Students should explain that, under the specified Poisson model, the probability of at least 3 complaint events in a full workday is approximately 32.33%. The parameter 2 is the expected number of complaint events within that window. Before students begin Material B, the instructor asks whether the workday probability can give the proportion of orders with a recorded complaint. Students explain how the statistical units and time windows differ. They also identify the order completion records and the linked contact records needed to calculate the proportion.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Table 3. Student Activities and Evidence for Feedback in the Complaint-Counting Example

Stage	Student tasks	Instructor review and follow-up	Materials submitted
Define the problem	Distinguish the probability associated with daily complaint counts from the proportion of orders with complaints.	Check what the question is about and which statistical unit applies	Statement of the problem and variables
Prepare the data	List the conditions in Material A and mark the order completion cohorts, follow-up windows, and repeat contacts in Material B.	Check waybill links and whether records cover each follow-up window	List of conditions and record assignments
Select a method	Explain the specified model and parameter in Material A. Define the numerator and denominator in Material B.	Check the rationale for using the specified model or defining the order-level proportion.	Specified model and parameter, and a justified definition of the proportion
Solve and check	Express the probability event and calculate its probability in Material A. Check counts and proportions in Material B.	Check event notation, inputs, and calculations. Revisit data preparation if record assignment or coverage is wrong.	Probability event, calculations, and checks
Interpret and revise	Specify what the results refer to, the dates they cover, and their limitations.	Review the rationale for the revision and decide whether to reduce prompts temporarily or return to data preparation	Revised conclusion and rationale
Reflect, summarize, and extend	Use the new records to decide independently whether the complete two-day proportions can be compared.	Check coverage and the conclusion. Return to Stage 2 for a coverage error or Stage 5 for an unsupported conclusion.	Individual response with reasoning

4.3. Using Contact Records to Revise the Order-Level Comparison

A student might reason that the one contact on June 2 and the two contacts on June 3 show that orders completed on June 3 were more likely to draw a complaint. The instructor identifies the confusion between contact intake dates and order completion cohorts. The instructor then removes identifying information from the relevant excerpt and asks GenAI to draft one question that directs the student to the waybill numbers. The instructor screens the draft for wording that gives away the cohort assignment, denominator, or conclusion, or treats repeat contacts as separate orders. The instructor might ask which order completion dates correspond to the two contacts received on June 3.

After checking the waybills, students find that the two contacts received on June 3 relate to orders in different order completion cohorts. The follow-up window for the June 1 cohort ends on June 3. The two contacts concerning A01 count as a single order in the order-level analysis. The follow-up window for the June 2 cohort ends on June 4. B01 and B02 are counted. The

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

contact concerning B03 falls outside the window. The observed proportions for the two cohorts are 1/4 and 2/8, respectively, both equal to 25%. These proportions describe only the orders in the constructed records. For customer service workload on June 3, both contacts are counted, including the repeat contact concerning A01.

The instructor checks whether students can explain the order completion cohort assignments, count each order with a complaint once, and exclude contacts outside the follow-up window. Once students justify these decisions, the instructor gives each student a new set of constructed records and asks for an individual response without an initial follow-up question. Four orders (C01 through C04) were completed on July 10. Six orders (D01 through D06) were completed on July 11. The supplied records cover all complaint-related contacts for these orders through the end of July 12. They show one contact for C01 on July 11 and one contact for D01 on July 12. The same two-day follow-up rule applies. Students decide whether the available records establish how the two complete follow-up proportions compare and explain their reasoning.

The July 10 cohort has complete coverage for July 11 and July 12. Its two-day proportion is 1/4. The July 11 cohort requires coverage for July 12 and July 13. Its observed proportion of 1/6 by the cutoff is a lower bound on its two-day proportion. Without the July 13 records, students cannot determine how the two complete follow-up proportions compare.

If a student treats July 13 as having zero complaints, the instructor returns the student to Stage 2 to check which dates the records cover. If a student identifies the incomplete follow-up window but still claims that the complete two-day proportions can be compared, the instructor returns the student to Stage 5 to revise the conclusion. The instructor keeps prompts reduced when the student can justify the limits of the comparison.

4.4. Identifying Data Requirements in the Open Exploration Task

In the open exploration task, students choose to study either workload associated with handling complaint-related contacts or the risk of complaints at the order level. They identify the data needed and explain how they would use them. Contact counts are a starting point for examining workload. Staffing decisions also require when contacts were received and how long they took to handle. To compare complaint risk at the order level, students check order completion cohorts, verify complete follow-up, and determine whether complaint records can be matched to orders. If students plan to develop a daily count model, they explain the relationship between operational exposure and the expected count and justify the proposed relationship. The fact that counts are nonnegative integers establishes only the possible values

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

of the variable. The mean, variance, and a time plot provide preliminary checks. A judgment about whether the Poisson model is appropriate also requires knowledge of the operations and how the records were generated. The open exploration task ends with a statement of data requirements and an analysis plan.

5. Discussion and Conditions for Implementation

5.1. Using Students' Written Work to Guide Classroom Feedback

If students submit only the final proportion, the instructor may be unable to tell whether they correctly assigned complaints to order completion cohorts or arrived at the same value by chance. Checkpoints during exploratory data analysis and feedback on drafts in statistics course projects give instructors opportunities to review students' interim decisions [10]. Interviews about graphing tasks trace how students adjust prompts and revise outputs, while research on statistical writing compares texts to examine how explanations appear in written responses [8,20]. Initial responses help instructors locate the step at which an error occurred. Explanations of revisions show whether students can account for their changes. For Material B, the instructor first checks whether students have mistaken the contact intake date for the order completion date. The instructor then checks whether they can use waybill numbers to reassign records and explain why they count repeated contacts only once per order and exclude contacts outside the follow-up window.

Instructors can prepare follow-up questions for common errors before class. Research on learning design prior to implementation has focused mainly on developing and drafting materials. Course examples also show that generated activities and assessments still need to be reviewed against course objectives and the teaching context [21,22]. In this example, the instructor checks the questions, records, and expected responses during preparation and selects prompts in class based on students' initial work. Students may distinguish the contact intake date from the order completion date and still record zero complaints for a day they have yet to observe. In that case, the follow-up question should address which dates the records cover.

5.2. Deciding When to Fade Prompts

The instructor needs to determine whether students can complete similar work independently when conditions change. The group that used Socratic dialogue performed better on measures of academic performance and reflection. Learning motivation showed no significant difference between groups [23]. In an art and design task, the group that selected its own scaffolds scored

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

higher on self-reported computational thinking and self-efficacy than the group that used GenAI alone [24]. A brief classroom comparison covering introductory AI topics at the university level found no consistent advantage for any level of prompting support [25]. The dialogue format, support for self-regulation, and intensity of prompting each vary a different aspect of instruction. Academic performance, self-reports, and motivation also address different questions. In the proposed instructional model, these distinctions inform a specific decision about whether students can independently explain the statistical reasoning behind their work on the current task.

After students explain their decisions about cohort assignment, repeat contacts, and record coverage, the instructor gives them the new July records in Section 4.3. Students complete the individual task without an initial follow-up question. This step uses a subsequent response to guide the fading of scaffolds and the transfer of responsibility to students [17]. If students identify the incomplete follow-up window and explain why the complete two-day proportions cannot yet be compared, the instructor can keep prompts reduced. If students treat an unobserved day as having zero complaints, the instructor restores support for checking record coverage. This response identifies a difficulty with follow-up completeness. Cohort assignment can be reviewed separately if the student's work reveals a difficulty with it.

5.3. Revisiting Earlier Stages of Inquiry and Conditions for Implementation

Comparing workload by contact intake date and comparing the proportion of orders with complaints by order completion cohort require different information from the same records. In statistical inquiry, new questions that arise during analysis can be addressed by revisiting the research question or the data [15,16]. In Material B, students first identify the statistical unit based on what they intend to compare and then check whether the records cover the follow-up windows. Errors in cohort assignment or the follow-up window require students to revisit the records even when the arithmetic is correct. If authentic data are used in class, students also need to examine how underreporting and missing records might affect the comparison.

Implementation requires time for instructors to review initial responses and provide feedback before students revise. If several responses show the same difficulty, the instructor can review a shared question and then use individual revisions to identify who still needs support. Simulated tutoring can draw on the state of the interaction to inform feedback [14]. In the proposed model, the instructor uses students' written work to choose a question and uses GenAI only to draft it. Students receive the question from the instructor and use their course tools to check their work. Before sharing student or company materials with an external tool, the

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

instructor should obtain permission and de-identify the materials. A classroom evaluation could examine whether instructors can review responses and provide timely feedback and whether students can explain their decisions under changed conditions after prompts are reduced.

6. Conclusion

The proposed model connects probability calculations under a specified model with decisions about what to count, which records to use, and how far a conclusion can reach. Its six stages specify what students submit and when they revisit earlier work. Instructors use GenAI only to draft follow-up questions and review those questions before sharing them.

In the individual task, a previously unseen set of records ends before one cohort completes its follow-up window. The student's explanation guides the instructor's decision to keep prompts reduced, restore support for checking record coverage, or ask for a revised conclusion. Classroom evaluation is needed to examine feasibility and student learning. A comparison with instructor-authored questions would help assess the added value of GenAI-assisted drafting.

Funding

This work was supported by the project "'GenAI Statistical Modeling'-Empowered Classroom Model Construction and Application Research for Probability and Statistics" (Project No. 2025JGX045). This work is also an output of the National First-class Undergraduate Major Construction Site, Major of Logistics Management, Nanning Normal University.

Declaration of Generative AI Use

The authors used generative AI to assist with English translation and language editing.

References

[1] GAISE College Report ASA Revision Committee. (2016). *Guidelines for assessment and instruction in statistics education college report 2016*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf

[2] Phadke, S., Beckman, M., & Lock Morgan, K. (2024). Examining the role of context in statistical literacy assessment. *Statistics Education Research Journal, 23*(1), Article 4. https://doi.org/10.52041/serj.v23i1.529

[3] Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education, 31*(2), 128–133. https://doi.org/10.1080/26939169.2023.2223609

[4] Xing, Y. (2024). Exploring the use of ChatGPT in learning and instructing statistics and data analytics. *Teaching Statistics, 46*(2), 95–104. https://doi.org/10.1111/test.12367

[5] Schwarz, J. (2025). The use of generative AI in statistical data analysis and its impact on

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

teaching statistics at universities of applied sciences. *Teaching Statistics, 47*(2), 118–128. https://doi.org/10.1111/test.12398

[6] Bien, J., & Mukherjee, G. (2025). Generative AI for Data Science 101: Coding without learning to code. *Journal of Statistics and Data Science Education, 33*(2), 129–142. https://doi.org/10.1080/26939169.2024.2432397

[7] Bray, R. L. (2025). A tutorial on teaching data analytics with generative AI. *INFORMS Journal on Applied Analytics, 55*(4), 319–343. https://doi.org/10.1287/inte.2023.0053

[8] Rao, V. N. V., Jeng, A., Drew, J., Kala, B., & Gongati, S. (2026). Students' statistical thinking when using generative AI: A descriptive case study. *Journal of Statistics and Data Science Education, 34*(1), 4–15. https://doi.org/10.1080/26939169.2025.2596684

[9] Al Labadi, L., & Ly, A. (2025). Enhancing statistics education through project-based learning (PBL) and the emergence of ChatGPT. *Teaching Statistics, 47*(3), 200–218. https://doi.org/10.1111/test.12405

[10] Al Labadi, L., Ly, A., & Kirusheva, S. (2026). Student perceptions of AI and virtual populations in a project-based experimental design course. *Teaching Statistics*. Advance online publication. https://doi.org/10.1002/test.70043

[11] Levine, R. A. (2026). Facilitating a collaborative relationship between generative AI and the statistics student. *The American Statistician, 80*(3), 360–371. https://doi.org/10.1080/00031305.2025.2608724

[12] Williams, K. Z., & Rudchenko, T. (2026). Transformative teaching in management statistics: Integrating real-world experiences and artificial intelligence. *Decision Sciences Journal of Innovative Education, 24*(1-2), Article e70019. https://doi.org/10.1111/dsji.70019

[13] Ugrinowitsch, C., & Libardi, C. A. (2026). From gatekeeper to architect: Operationalizing AI as a cognitive partner in higher education. *Frontiers in Education, 11*, Article 1880704. https://doi.org/10.3389/feduc.2026.1880704

[14] Zhang, J. (2026). Scaffolding probabilistic reasoning in civil engineering education: Integrating AI tutoring with simulation-based learning. *Education Sciences, 16*(1), Article 103. https://doi.org/10.3390/educsci16010103

[15] Pittard, M., Saidi, R., Bargagliotti, A., & Malone, C. (n.d.). *Recommendation 1*. Guidelines for Assessment and Instruction in Statistics College Report. https://amstat.quarto.pub/college-gaise/recommendations/recommendation-01.html

[16] Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

[17] van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review, 22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

[18] Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

[19] NIST/SEMATECH. (n.d.). Poisson distribution. In *e-Handbook of statistical methods*. Retrieved November 11, 2026, from https://itl.nist.gov/div898/handbook/eda/section3/eda366j.htm

[20] DeLuca, L. S., Reinhart, A., Weinberg, G., Laudenbach, M., Miller, S., & Brown, D. W. (2025). Developing students' statistical expertise through writing in the age of AI. *Journal of Statistics and Data Science Education, 33*(3), 266–278. https://doi.org/10.1080/26939169.2025.2497547

[21] Alkhasawneh, S. N., Lampropoulos, G., & Hernández-Leo, D. (2025). Generative AI in learning design: A systematic review. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, &

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

S. Isotani (Eds.), *Artificial intelligence in education: 26th international conference, AIED 2025, Palermo, Italy, July 22–26, 2025, proceedings, part IV* (pp. 17–30). Springer. https://doi.org/10.1007/978-3-031-98459-4_2

[22] Pereira, E., Nsair, S., Pereira, L. R., & Grant, K. (2024). Constructive alignment in a graduate-level project management course: An innovative framework using large language models. *International Journal of Educational Technology in Higher Education, 21*, Article 25. https://doi.org/10.1186/s41239-024-00457-2

[23] Xi, L., Zhang, Y., & Wang, Q. (2026). Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education. *Computers & Education, 241*, Article 105494. https://doi.org/10.1016/j.compedu.2025.105494

[24] Xia, Q., Yang, Y., Wang, W., & Yin, H. (2026). Promoting interdisciplinary learning with generative AI through self-regulated scaffolding. *Journal of Computer Assisted Learning, 42*(2), Article e70214. https://doi.org/10.1002/jcal.70214

[25] Melanou, C., & Beege, M. (2026). Scaffolding generative AI as a tutor: A quasi-experimental study of learning outcomes and motivational, cognitive and metacognitive processes. *Education Sciences, 16*(4), Article 651. https://doi.org/10.3390/educsci16040651

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.